CERTIFICATE OF DELIVERY			
I hereby certify that this correspondence is being hand-delivered to the U.S. Patent and Trademark Office, Art Unit 1655			
Typed or Printed Name	Donna Macedo		
Signature	Donna Macedo	**Date**	3/17/03

SUPPLEMENTAL DECLARATION UNDER 37 C.F.R. § 1.132 Address to: Assistant Commissioner for Patents Washington, D.C. 20231		
	Attorney Docket Confirmation No.	P 4941
	First Named Inventor	Chenchik et al.
	Application Number	09/440,829
	Filing Date	November 15, 1999
	Group Art Unit	1634
	Examiner Name	Forman, B.
	Title	*Long Oligonucleotide Arrays*

Rule 132 Declaration

I, Alexander Munishkin, am a co-inventor on the above cited patent application. I received my undergraduate academic training at Samara State University, and was awarded my BS degree in biology in 1983. My graduate work continued at that same institution, where I received by MS degree in biology in 1985. I continued with my studies at Moscow University, and earned my PhD degree in molecular biology in 1991. After completing my formal education, I undertook my post-doctoral position at University of Chicago from 1992-1998. I joined Clontech's Array research group in 1998, starting as Scientist 1. Since that time, I have advanced steadily to my current position as Scientist 3, Group Leader.

Prior Literature and Microarrays

In 1999, the expectation of the research community was that there would be little or no increase in sensitivity of nucleic acid mixtures spotted on a microarray if their nucleic acid strand length was increased. Schwille et al. 1996 (**Exhibit A**) showed that for 17mers, 19mers, 23mers, 29mers, 30mers, and 37mers that the amount of bound

probe, a direct indicator of sensitivity, was not related to the length of the nucleic acid chains spotted on the array (see page 10,183, under Materials and Methods, and Table 1, at page 10,187). This data has been collaborate many times since, as described below in the section on experimentation.

By 1999, the 25mer range nucleic acid strand length produced by standard synthetic methods was well established for use on microarrays. These methods of DNA strand synthesis and the 25mer range continues to be the basis for the Affymetrix array products.

The other established, reasonably priced method to produce DNA for microarray applications in 1999 was cDNA technology. By example, in Clontech's product lines at that time, 200 to 2,000mer cDNA strands as products of PCR amplification as routine in array product production. By 1999, other manufactures, such as Incyte and Synteni, were also starting to produce cDNA microarray products in this range.

Both studies in the published literature and practical experience with manufactured arrays indicated that the sensitivity of the cDNA arrays, which range upwards of 200mers, were comparable to the established 25mer arrays. These developments further entrenched the research community's assumption that an increased length of the NA spot did not increase sensitivity.

60-120mers of any reasonable quality cannot be produced using either the standard array synthetic methods still employed commercially by Affymetrix, or cDNA using PCR. While available commercially in 1999, materials of the 60-120mers were impractically priced for the many strands necessary for each spot provided in microarrays. The cost of 60-120mers was typically $220 for each DNA strand. 60-120mers were available from production house employing the existing, published literature on phosphoramidite chemistry.

The literature and scientific community understanding in 1999, as well as cost considerations, would have strongly discouraged a researcher from investigating the use of 60-120mers on arrays. Not surprisingly, there was no reported experimentation into that possibility through this period.

In our laboratory, however, we took the unprecedented step of investigating the 60-120mers for use on microarrays. The use of these materials was, in part,

serendipitous because we had in-house expertise on the long-oligo synthesis process and could have them produced at a reasonable cost.

Our motivation in undertaking this study was not to find a more sensitive length. In fact, since laboratory was convinced, along with the rest of the research community, that increasing the length of the spotted nucleic acid strands would not improve the sensitivity of the microarray.

We undertook studies of specific sub-sequences in the 60-100mer range for a completely different purpose, that is to identify which of these partial gene sequences within specific genes of interest would bind with the greatest specificity to the sample materials. These constructs were produced merely as experimental materials, and with no intent that they would ever be used as a substitute for our standard cDNA products.

Thus, our laboratory was the first to undertake and report on an investigation of 60-120mers on microarrays. The surprising highly advantageous results have been demonstrated both in our laboratories, and now have been reported by other laboratories, as described below.

Experimental Data

I completed a number of experiments comparing progressively longer gene sequence fragments in a glass based microarray format. These experiments demonstrated the advantages in sensitivity of the inventive 60-120mers over the 25mers and cDNAs used previously on microarrays. In doing these experiments, I came to recognize this unexpected range of lengths has many advantageous qualities.

This recognition of the special advantages of 60-120mers on microarrays has, since the time of the filing of the application, been appreciated and adopted by the majority of industry leaders in the microarray field, as well as several of the University core facilities that have been organized in the last few years.

Concerning the data which I am reporting in this declaration, Exhibit B is a graph providing the intensity levels measured of varying nucleic acid strand lengths of spotted material on the surface of a microarray for four different gene examples. It includes my original work prior to the filling of the application, as well as additional data points developed at the suggestion of the Examiner in our case.

The intensity levels quantified by the data represent fluorescent units in the sample nucleic acid molecules. The intensity level registered by the fluorescent scanner reading of the array after application of the sample nucleic acids to the microarray is indicative of the number of target molecules which bind to the array. Therefore, an increase in fluorescent intensity is directly correlated to the sensitivity of the assay.

Note that as shown in **Exhibit B**, the overall intensity of each gene varies one to another. However, there is an omnibus quality of the inventive technology that the 60-120mer spotted material for each gene provides a highly advantageous sensitivity as compared to both the lower ranges and higher ranges used by the scientific community prior to our invention.

The graph shown in **Exhibit B** provides the results of comparing the sensitivity of oligos of various lengths for specific genes, in this case mitochondria glycerol-3-phosphate dehydrogenase 2, creatine kinase, pyruvate dehydrogenase complex, and tubulin. I conducted further testing to provide additional data points from my original work, providing the following points of comparison: 20mer, 30mer, 35mer, 40mer, 50mer, 60mer, 80mer, 100mer, 120mer, and a 260-528mers. The last bar within each gene notation is for a cDNA synthesized sequence in the 260-528mer range, as noted. The unexpected sensitivity in the 60-120mer range is supported by this data.

To further understand the trends for sensitivity vs. sequence length, I have compared the sensitivity for 80mer to the gene's corresponding cDNA sequences, ranging from 260mers to 528mers. The graph in **Exhibit C** displays the 80mer sensitivity for each gene as 100% in the first bar, and the various cDNA length available for each gene as a comparative percentage. The trend is clearly demonstrated that 200-700mers have typically only half or less of the sensitivity of 80mers.

While the intervening data between 120mers and 260-700mers is not available for the technical reasons above, it is a likely that the total graph for 20mers to 700mers shows a generally flat profile, with a slightly higher plateau from 200mers onwards as compared to the level of 40 or 50mers. However, highly unexpectedly given those prior know set points, is the peak in the 60-120mer range.

From my laboratory experience with 250 other genes (data not included here), in the 400mer-850mer range I have never seen the specificity and intensity at a level higher than the 50mer for that same gene.

There are some variation from gene to gene as to the strand length around the 60-120mer length where the highest peak in sensitivity we discovered occurs. However, as a matter of averages, 60-120mers consistently provide a significant increase in sensitivity over previously employed lengths. The smaller and larger maximum peak ranges are for practical purposes comparable as they average from the 60-120mers.

Given the averages of the large number of genes represented on a microarray, this length range is the best choice for gaining the best advantage of the inventive length, as born out in data from other laboratories, and the commercial retooling of many production facilities from the prior established ranges to 60-120mers, now generally termed in the industry as "longmers" or "long oligos".

Experimental Data from Other Laboratories

Data published by other laboratories further substantiates the unexpected value of employing longmers (typically 60-120mers) in microarrays. These announcements were made hand in hand with launch of new products which employ these ranges. However, the production of longmers can still be expensive. These cost have been brought down with greater facility in optimizing the production methods which has been accomplished in academia, and it is not surprising that these other laboratories are academic or have close academic association. It is impressive that the advantages of the 60-120mers is so substantial that these groups have undertaken this retooling of their approaches.

The Qiagen Operon Product Guide 2002 (attached as **Exhibit D**) provides a comparison of oligo proves of different length within 10 different yeast genes at 35mer, 50mer, 70mer and 90mer in Figure 2.2 at page 61. This data closely parallels that of my own research above. The exact position of the highest point of the peak can vary from gene to gene, but in all cases, shows an ascendance of sensitivity, and in at least half the cases show a drop in sensitivity at longer nucleic acid lengths. This lesser level of drop in sensitivity from my data is logically due to the highest level being reported at 90mer in the Qiagen data. From prior data on cDNAs, it would be expect that this drop in

sensitivity would continue to drop if reported in additional, longer examples. Note also Figure 2.4 at page 62, which shows that 70mers are superior in reliability that 50mers.

The MWG BioTech technical article (attached as **Exhibit E**) also evaluated varying length spots for sensitivity for 12 distinct yeast genes. In this case, 40mer, 50mer and 70mer lengths were evaluated. While in 2 cases, some drop in sensitivity appears at 70mer, showing a shift of the top of the peak in these genes to the smaller end, the remaining 10 cases show a peak at 70mers. All these findings are consistence with our recognized advantageous length.

Commercial Recognition of Advantageous Range

Beyond those companies which have published data demonstrating the advantage in sensitivity of longmers in microarrays are many others who have announced such advantages in concert with the launch of products employing materials in the 60-120mer range.

In the BioArray News (attached as **Exhibit F**), several company announcements of long oligos microarray product launches are reported. Dan Shoemaker of Rosetta Inpharmatics announced "..we found that 60-mers were ideal", reporting that this range had the highest specificity and sensitivity of all lengths. Agilent presented data demonstrating that its protocol using 60mers provided results that were more reproducible, sensitive and accurate than other protocols.

Additionally, core faculties at academic institutions have reported that they are using long oligos to spot their own arrays. GlaxoSmithKline's Pritit Hegde reports on 30 and 40 mers vs. 50-70mers that the latter had "a far lower rate of random hybridization". Harvard Medical School lab found in its own studies that 70mers delivered the best results, and has since employed oligos of those lengths in their arrays.

Summary

In conclusion, in 1999 there was no motivation for researcher to employ 60-120mers in microarrays. Based on the research available at that time, it was widely assumed that there was no advantage in sensitivity to longmers. As a results, there was little motivation to experiment in that area. Additionally, the methods of producing

nucleic acid strands in the 60-120mer range although well known, was sufficiently expensive and time consuming that researchers were further discourage from perusing this investigation. Ours was the only laboratory to investigate and report this possibility.

Since our introducing 60-120mers to the market, the industry has in large part moved to this preferred range. As above, they have reported the unexpectedly advantages results of using the 60-120mers on arrays.

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code; and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

March 17, 2003

Signature: A. Munishkin

Alexander Munishkin

EXHIBIT A

Quantitative Hybridization Kinetics of DNA Probes to RNA in Solution Followed by Diffusional Fluorescence Correlation Analysis†

Petra Schwille,* Frank Oehlenschläger,‡ and Nils G. Walter§

Max-Planck-Institute for Biophysical Chemistry, Department of Biochemical Kinetics, Am Fassberg, D-37077 Göttingen, Germany

Received March 1, 1996; Revised Manuscript Received May 31, 1996®

ABSTRACT: Binding kinetics in solution of six *N,N,N′,N′*-tetramethyl-5-carboxyrhodamine-labeled oligodeoxyribonucleotide probes to a 101mer target RNA comprising the primer binding site for HIV-1 reverse transcriptase were characterized using fluorescence correlation spectroscopy (FCS). FCS allows a sensitive, non-radioactive real time observation of hybridization of probes to the RNA target in the buffer of choice without separation of free and bound probe. The binding process could directly be monitored by the change in translational diffusion time of the 17mer to 37mer DNA probe upon specific hybridization with the larger RNA target. The characteristic diffusion time through a laser-illuminated open volume element with 0.5 μm in diameter increased from 0.13−0.2 ms (free) to 0.37−0.50 ms (bound), depending on the probe. Hybridization was approximated by biphasic irreversible second-order reaction kinetics, yielding first-phase association rate constants between 3×10^4 and 1.5×10^6 M^{-1} s^{-1} for the different probes. These varying initial rates reflected the secondary structures of probes and target sites, being consistent with a hypothetical binding pathway starting from loop−loop interactions in a kissing complex, and completion of hybridization requiring an additional interaction involving single-stranded regions of both probe and target. FCS thus permits rapid screening for suitable antisense nucleic acids directed against an important target like HIV-1 RNA with low consumption of probes and target.

Hybridization of nucleic acids to their complementary sequences is a fundamental process in molecular biology. It plays a major role in replication, transcription, and translation, where specific recognition of nucleic acid sequences by complementary strands is essential for propagation of information content. In most of these processes, RNA participates as the naturally occurring single-stranded nucleic acid form, ready to hybridize. Competing with hybridization to another single-stranded molecule, formation of secondary structure via intramolecular hydrogen bonds can occur. The secondary structure of RNA is also involved in other processes like binding of specific proteins, hydrolysis within the cellular environment, or transcription and translation control (Ma et al., 1994; Yang et al., 1995; Varani, 1995).

In the case of naturally occurring antisense RNAs, hybridization plays a negative feedback role. These molecules specifically bind to their complementary sequences and thereby block functionality of sense RNA (Simons, 1988; Wagner & Simons, 1994). This has been used to design artificial antisense RNAs to down-regulate target gene expression (Inouye, 1988; van der Krol et al., 1988; Wagner, 1994). Both RNA and DNA probes are currently employed to suppress viral replication, a method that might become a therapeutic tool to particularly fight pathogenic retroviruses (Crooke, 1992; Dropulic & Jeang, 1994). With viruses such as HIV-1, the viral RNA is simultaneously a target for hybridization of the replication primer (typically a host tRNA) and the therapeutic antisense nucleic acid, both being in competition with secondary structure formation of their target sites (Lima et al., 1992; Isel et al., 1995). Consequently, hybridization between complementary strands is complex and initiates at loops or bulges within the secondary structure, followed by rapid zippering leading to fully double-stranded hybrid (Wagner & Simons, 1994; Hjalt & Wagner, 1995). It is therefore not surprising that the performance of a particular antisense nucleic acid is often not predictable within a host cell, where both target and antisense strand might be inaccessible due to higher order structures and complexation with proteins or hybridization might simply be unfavorable because of ionic conditions and low concentrations.

A better understanding of RNA hybridization to complementary strands in solution could provide deeper insights into the described fundamental biological and technological processes. Thus, it becomes necessary to perform kinetic analyses of nucleic acid hybridization. Classically, these analyses have been performed to understand gene structure and function, especially genome complexity and gene copy number (Britten & Kohne, 1968; Young & Anderson, 1985). The basic requirement for a quantitative study on nucleic acid hybridization in solution is to separately monitor paired and unpaired strands. In the past, this has been achieved using physical methods like absorbance spectroscopy (hypochromicity or circular dichroism; Bush, 1974), calorimetry (Breslauer, 1986), or nuclear magnetic resonance (Patel et al., 1982). Generally, these techniques require quite considerable amounts of nucleic acids in the microgram to milligram range. Radioactive labeling allows detection of

† This work was supported by Grant No. 0310739 from the German Ministry for Education, Science, Research, and Technology. Financial support by EVOTEC BioSystems GmbH, Hamburg, to P.S. and F.O. is gratefully acknowledged.

* Author to whom correspondence should be addressed. Tel: +49-551-201-1436. FAX: +49-551-201-1435.

‡ This author should be regarded as also having first author status.

§ Present address: University of Vermont, Department of Microbiology and Molecular Genetics, Burlington, VT 05405.

® Abstract published in *Advance ACS Abstracts*, July 15, 1996.

S0006-2960(96)00517-X CCC: $12.00

minute amounts of nucleic acids and has been used for direct analysis of solution hybridization on non-denaturing gels (Kumazawa et al., 1992) or by chromatographic methods (Dewanjee et al., 1994) and for enzymatic assays like resistance to nuclease S1 (Bishop et al., 1974) or RNase H (Zarrinkar & Williamson, 1994). With these isotopic assays, physical separation of hybridized and unhybridized strands is required, e.g., by precipitation, solid phase capturing, electrophoresis, or chromatography. This makes true solution-phase measurements impossible.

Recently, sensitive fluorescence measurements have been used to directly monitor nucleic acid hybridization in solution. One approach uses a fluorophore on the 5′ end of one strand and a quenching dye on the 3′ end of the complementary strand. Hybridization is then monitored by decreasing fluorescence of the donor and increasing fluorescence of the acceptor due to starting energy transfer (Morrison & Stols, 1993). This technique requires two fluorescent labels at different sites and so far has been limited to hybridization studies of complementary DNAs forming a blunt-ended hybrid. In an analogous approach, the same strand is labeled with a donor on the 3′ end and an acceptor on the 5′ end and energy transfer decreases after hybridization (Parkhurst & Parkhurst, 1995). With certain fluorophores like pyrene, the detection of hybridization to a complementary strand is possible due to altered quenching effects of base-paired nucleobases on the dye. Either DNA−DNA (Manoharan et al., 1995) or RNA−RNA hybridization in solution (Li et al., 1995) can thus be monitored, but typically quite high (micromolar) concentrations of the labeled strand are required.

Fluorescence correlation spectroscopy (FCS)[1] is a technique developed to study dynamic processes of fluorescent molecules that give rise to fluorescence fluctuations (Magde et al., 1972, 1974; Elson & Magde, 1974; Ehrenberg & Rigler, 1974; Koppel, 1974). Since its introduction, the technique has found a broad range of applications, like measurement of diffusion constants, chemical kinetic rate constants, and molecular weights [for review, see Thompson (1991)]. Recently improved setups use an epi-illuminated microscope with strong focusing of the exciting laser beam and a small pinhole with an avalanche diode for detection, e.g., to analyze translational diffusion in dilute solutions (Rigler et al., 1992, 1993). Kinjo and Rigler (1994) were thus able to follow the binding of a fluorescently labeled 18mer DNA primer at a concentration of 50 nM to a 7.5 kb DNA containing the complementary sequence by monitoring the slowing down of primer diffusion through the laser beam.

To understand hybridization to RNA strands, we have been interested in hybridization kinetics of DNA probes to RNAs. Here, FCS seemed to be an appropriate tool, since it allows direct observation of hybridization without physical separation of strands, but with high sensitivity and requiring only the DNA strand to be labeled with a single, freely eligible fluorophore. Since many biologically relevant RNAs (like tRNAs or ribozymes) are often between 70 and 700 bases in length and since diffusion times (being inversely related to diffusion coefficients) are in first approximation proportional to the third root of the molecular weight of the diffusing species (according to the Stokes−Einstein relation), the increase in diffusion time of the labeled probe upon hybridization can be expected to be low and quantitative values of hybridized fractions difficult to extract. In the present work, we therefore used an artificial short-chained RNA comprising the replicative primer binding site of HIV-1 to rigorously prove that FCS can measure quantitative kinetic constants for this kind of hybridization targets. The 101mer RNA folds into a secondary structure with two stem−loop domains (Figure 1) and has been used in our laboratory as template for *in vitro* replication studies with reverse transcriptases (Pop, 1995; Gebinoga & Oehlenschläger, 1996). We designed six *N,N,N′,N′*-tetramethyl-5-carboxyrhodamine (TMR)-labeled DNA probes with equal calculated melting points against different regions of the target (Figure 1) and were able to directly monitor the increase in their diffusion times upon binding in solution by a shift in the autocorrelated fluorescence signal. Using appropriate controls, quantitative data for the ratio of bound to unbound species at a total concentration of 10 nM could be extracted and compared to values obtained by a non-radioactive primer extension assay using the same fluorescent probes. Thus, it could be shown that DNA−RNA hybridization kinetics as a function of target and probe secondary structure can directly and sensitively be followed using FCS.

MATERIALS AND METHODS

Materials. Target α-1 RNA is a 101-nucleotide *in vitro* transcript of the plasmid HP18α-1, linearized with *Hin*dIII (Pop, 1995), its concentration being determined by the assumption that 1 OD_{260} equals 40 μg/mL. It shows a secondary structure with some double-stranded regions (Figure 1). Using the Vienna RNA package computer program (Hofacker et al., 1994), a denaturation temperature of about 70 °C was calculated. The six DNA probes HS1 to HS6 are labeled with the 5-isomer of TMR at their 5′ end via an aminohexyllinker (Figure 2) and were purchased in HPLC-pure quality from NAPS (Göttingen, Germany). Their purity was again controlled by HPLC (monitoring absorbances at 260 and 554 nm), their concentration determined taking into account, that the TMR label contributes to the absorbance at 260 nm (with $A_{260}/A_{554} = 0.49$) and the degree of substitution (DOS) confirmed to be one label per molecule using the equation $DOS = [(10N/86)A_{554}]/[A_{260} - (0.49 \times A_{554})]$ (with N the number of bases in the probe). Sequences were as follows: 19mer HS1, 5′-TMR-d(GACATTGTTCGTCGGCCGC); 29mer HS2, 5′-TMR-d(CATCAATGTCAATAAGGTGACATTGTTCG); 37mer HS3, 5′-TMR-d(TGCTAGAGATCTCTAAGTTATAACACATCAATGTCAA); 30mer HS4, 5′-TMR-d(GGCGCCACTGCTAGAGATCTCTAAGTTATA); 17mer HS5, 5′-TMR-d(GTCCCTGTTCGGGCGCC); 23mer HS6, 5′-TMR-d(AGCTTCCCTTTCGCTTTCA GGTC). The probes were chosen such that each probe's complex with its cDNA would melt in hybridization buffer at about 77 °C, suggesting uniform thermodynamic parameters for the RNA−DNA hybrids as well. HS1X−HS6X are the corresponding unlabeled probes and GSHS1−GSHS6 are the length-matched cDNA strands of HS1−HS6, respectively, all being synthesized on a Milligene Expedite Synthesizer. HIV-1

[1] Abbreviations: FCS, fluorescence correlation spectroscopy; kb, kilo bases; TMR, *N,N,N′,N′*-tetramethyl-5-carboxyrhodamine; HIV, human immunodefficiency virus; HPLC, high-performance liquid chromatography; DOS, degree of substitution; bp, base pairs; PAGE, polyacrylamide gel electrophoresis; PACE, polyacrylamide capillary electrophoresis.

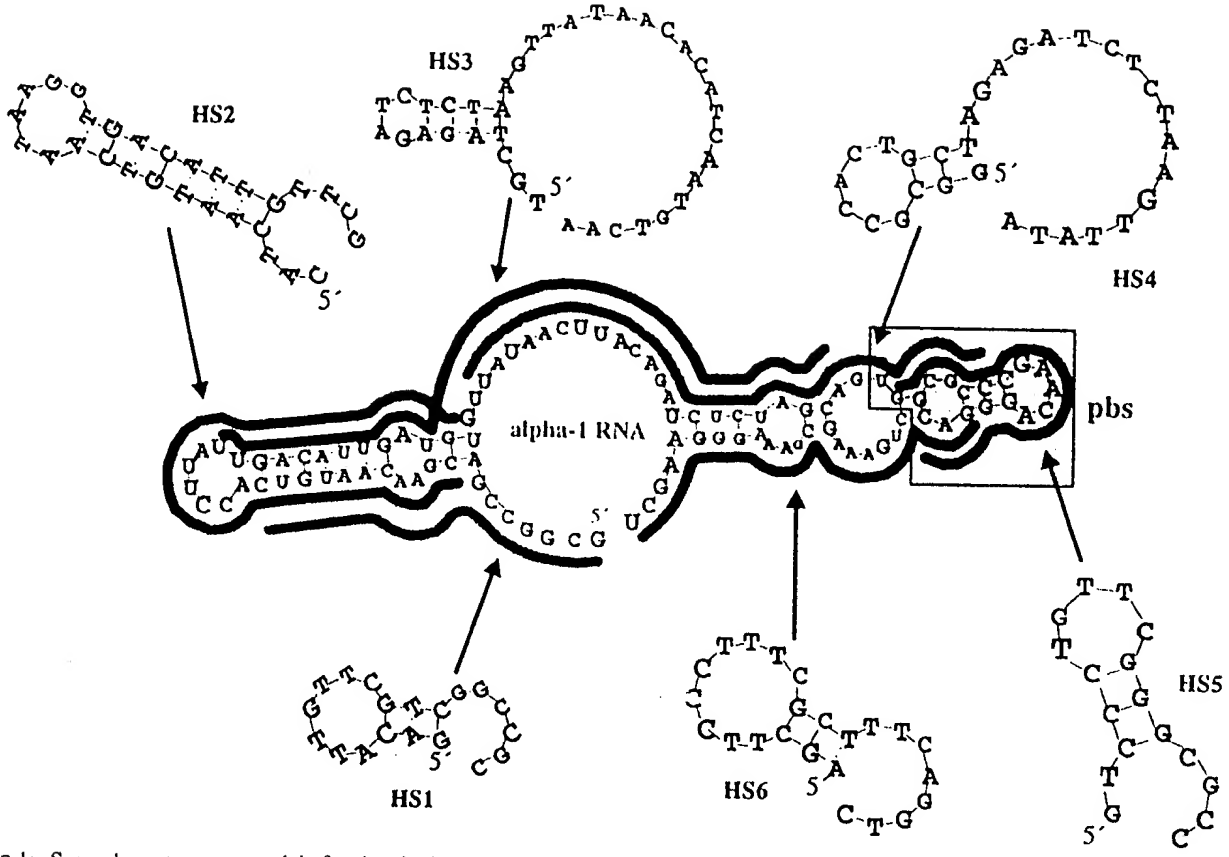

FIGURE 1: Secondary structure models for the six DNA probes HS1 to HS6 and target α-1 RNA. The 5′ TMR-labeled probes are designed to hybridize with different target sites, represented by shaded lines. The primer binding site (pbs) of HIV-1 reverse transcriptase is highlighted by a shaded bar.

reverse transcriptase was a grateful donation from Dr. Magda Pop and purified from an overexpressing *Escherichia coli* strain as described (Müller et al., 1989). Sonicated salmon sperm DNA was from Stratagene (Heidelberg, Germany). dNTPs were obtained from Pharmacia (Freiburg, Germany), while TMR-labeled UTP was custom-made by NAPS (Göttingen, Germany).

Hybridization Protocols. For kinetic analysis, α-1 RNA was dissolved in water to 1 μM, heated at 75 °C for 2 min to ensure complete denaturation, and allowed to cool to room temperature for 15 min. This stock solution was used to set up solution A with typically 100 nM α-1 RNA in 60 mM Tris-HCl, pH 8.2, 10 mM $MgCl_2$, 10 mM KCl, 2.5 mM DTT, 2 mM spermidine, and 10 μg of sonicated salmon sperm DNA/mL. Solution B typically contained 20 nM HS1−HS6 (60 nM in the case of HS3) in the same buffer excluding RNA. Both solutions were equilibrated separately at 40 °C for 30 min. Hybridization was initiated by mixing equal volumes of solutions A and B (typically each 50 μL) at 40 °C. 30 μL aliquots were continuously analyzed in an open sample carrier at 40 °C by FCS, being exchanged after 5 min to limit deviations due to sample evaporation, adsorption, or bleaching.

To measure a maximum value for hybridization extent, solutions A and B described above were mixed and then heated at 75 °C for 2 min. The mixture was cooled to room temperature over 15 min and then incubated at 40 °C for 15 min before FCS analysis. To study dissociation, an excess of 1 μM unlabeled probe was added to the obtained hybrid and the diffusion time of the TMR-labeled probe was monitored over 2 h.

For hybridization of corresponding cDNA strands with HS1−HS6, solution A contained 1 μM GSHS1−GSHS6 instead of 100 nM α-1 target RNA, and both solutions A and B were first mixed, then denatured, and cooled down as described above.

To hybridize TMR-labeled α-1 RNA with excess unlabeled probe, solution A described above contained 20 nM TMR-labeled α-1 RNA, while solution B included 1 μM unlabeled HS1X−HS6X. Both solutions were again mixed prior to denaturation and cooled down as described above.

To measure the diffusion time of α-1 RNA in dependence of initial RNA concentration, 125 nM TMR-labeled α-1 RNA was mixed with 1.25 μM unlabeled α-1 RNA and diluted to give total RNA concentrations of 1.38 μM, 550 nM, 275 nM, and 138 nM in water. An additional solution contained 40 nM TMR-labeled α-1 RNA. These solutions were heated at 75 °C for 2 min, cooled to room temperature over 15 min, diluted to a final concentration of 138 nM RNA (40 nM for the fifth solution) in hybridization buffer (60 mM Tris-HCl, pH 8.2, 10 mM $MgCl_2$, 10 mM KCl, 2.5 mM DTT, 2 mM spermidine, and 10 μg of sonicated salmon sperm DNA/mL), and analyzed by FCS.

FCS Measurement and Extraction of Diffusion Times and Hybrid Fractions. Fluorescence correlation spectroscopy is a special case of fluctuation correlation spectroscopy, where temporal fluctuations in a sample of laser-excited fluorescent molecules are self-correlated to obtain information about the processes leading to fluorescence fluctuations. These underlying processes may be photophysical transitions, shifts in wavelength, changes in quantum yield, or simply concentration fluctuations by thermal motion (diffusion) of the

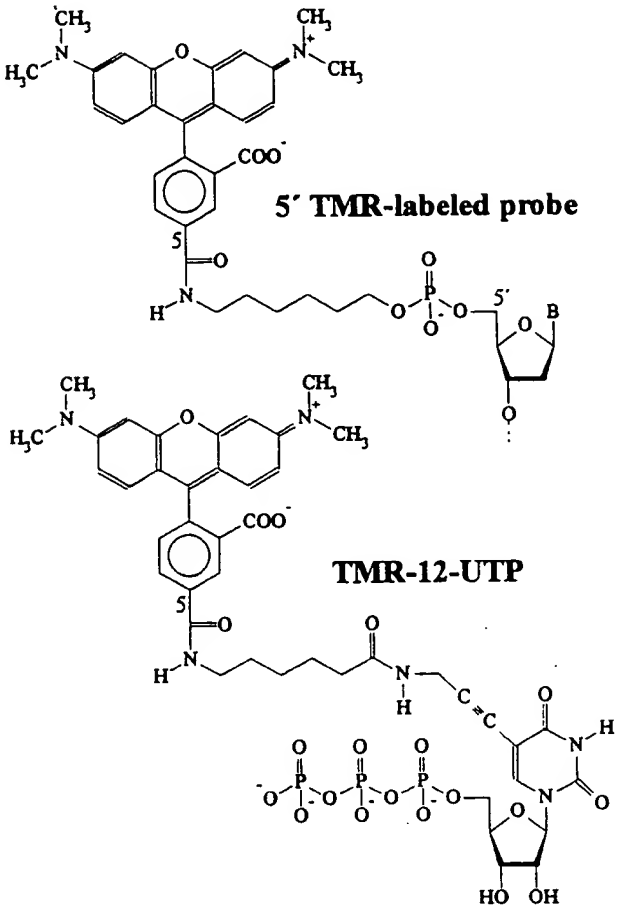

FIGURE 2: Molecular structures of the N,N,N′,N′-tetramethyl-5-carboxyrhodamine (TMR) labels used in this study for fluorescent detection. DNA probes were 5′ end-labeled with TMR-succinimidyl ester via an aminohexyl linker, while RNA was internally labeled by transcription in the presence of TMR-12-UTP.

fluorophores. In solutions with diffusing species, both the magnitude $G(0)$ and the rate and shape of the temporal decay of the autocorrelation function $G(t)$ have previously been used to detect concentrations and characterize molecular aggregation (Palmer & Thompson, 1989; Thompson, 1991). The temporal decay of $G(t)$ allows extraction of the characteristic time for diffusion of the fluorophores, which may change upon interaction with non-fluorescent molecules. This latter principle was used earlier to analyze binding of fluorescently labeled antigens or antibodies to latex particles (Briggs et al., 1981) or of DNA probes to a DNA target (Kinjo & Rigler, 1994) and was exploited in the present study for analysis of DNA–RNA hybridization.

Figure 3 describes our experimental setup. The 514 nm line of an argon ion laser (Lexel 85, power 0.2 mW) epi-illuminates a Zeiss water immersion 63 × 1.2 microscope objective without any prefocusing system. The sample droplet (30 μL) is placed into a gold-covered, chemically inert open sample carrier (Walter & Strunk, 1994) thermostated at 40 °C, and the objective surface is directly lowered onto the solution. Evaporation is minimized by close contact between sample carrier and objective, and adsorption and bleaching effects are reduced by exchange of the sample droplet after 5 min against solution separately incubated at 40 °C in a closed, light-shielded tube. The wavelength-shifted fluorescence light in opposite direction now traverses the dichroic mirror, passing a 565 DF 50 bandpass filter (Omega Optics) to suppress background light such as Raman scattering or laser reflections. The 50 μm diameter pinhole in the image plane defines the z-dimension of the analyzed sample volume and is imaged 1:1 onto the detector surface of an avalache photodiode (EG&G SPCM-200). The photocount signal was autocorrelated over 1 min (30 s for the first measurement after hybridization start) quasi-online by a digital signal correlator card (ALV-5000, Fa. Peters, Langen, Germany).

The autocorrelation function $G_1(t)$ for fluctuations in a diffusional system with a single sort of fluorescent particles depends on the average number of fluorophores N in the illuminated volume element of the sample (i.e., their concentration), the average translational diffusion time τ_{diff} (given by the xy-radius r of the volume element and the diffusion coefficient D to $\tau_{diff} = r^2/4D$), and the structure parameter of the volume element r/z (radius divided by half of the length), which is constant for a defined setup, in our case 0.2. Using the pinhole as optical field diaphragm (Qian & Elson, 1991), the three-dimensional shape of the illuminated detection volume element can be approximated as Gaussian in all directions (Rigler et al., 1993). This defines $G_1(t)$ to be (Thompson, 1991; Rigler et al., 1993):

$$G_1(t) = \frac{1}{N}\frac{1}{1+\frac{t}{\tau_{diff}}}\frac{1}{\sqrt{1+\left(\frac{r}{z}\right)^2\frac{t}{\tau_{diff}}}} \quad (1)$$

In the case of singlet–triplet transitions of the fluorophores and with T being the average fraction of dye molecules in triplet state with a relaxation time τ_{tr}, this changes to (Widengren et al., 1994, 1995):

$$G_{1,T}(t) = (1 - T + Te^{-t/\tau_{tr}})G_1(t) \quad (2)$$

The principle of hybridization detection is based on the sensitivity of FCS to changes in the average translational diffusion time. For a system of M diffusing species labeled with fluorophores of comparable triplet decay times, and with Y_i being their fractions ($\sum Y_i = 1$), the general autocorrelation function is given by:

$$G_{M,T}(t) = \frac{(1 - T + Te^{-t/\tau_{tr}})}{N}\sum_{i=1}^{M}\frac{Y_i}{1+t/\tau_i}\frac{1}{\sqrt{1+(r/z)^2\,t/\tau_i}} \quad (3)$$

If the diffusion times τ_i of the different components are known, the fractions can be determined in a sample droplet by mathematical rather than physical separation. Upon hybridization of a labeled DNA probe to its RNA target, a more slowly diffusing complex forms. The three to four times larger hybrid needs approximately twice as long to traverse the laser-illuminated volume element and remains stable throughout this diffusion time (in the range of ms), since dissociation is orders of magnitude slower. Theoretically, an $M = 2$ system is obtained, and the fraction of bound probe Y_2 increases over hybridization time. Eventual changes in triplet decay times τ_{tr} generally do not interfere with measured diffusion times, since for rhodamine dyes in water τ_{tr} are typically 2 orders of magnitude smaller and can easily be separated (Widengren et al., 1994, 1995). The dependence of triplet fraction and fluorescence quenching on

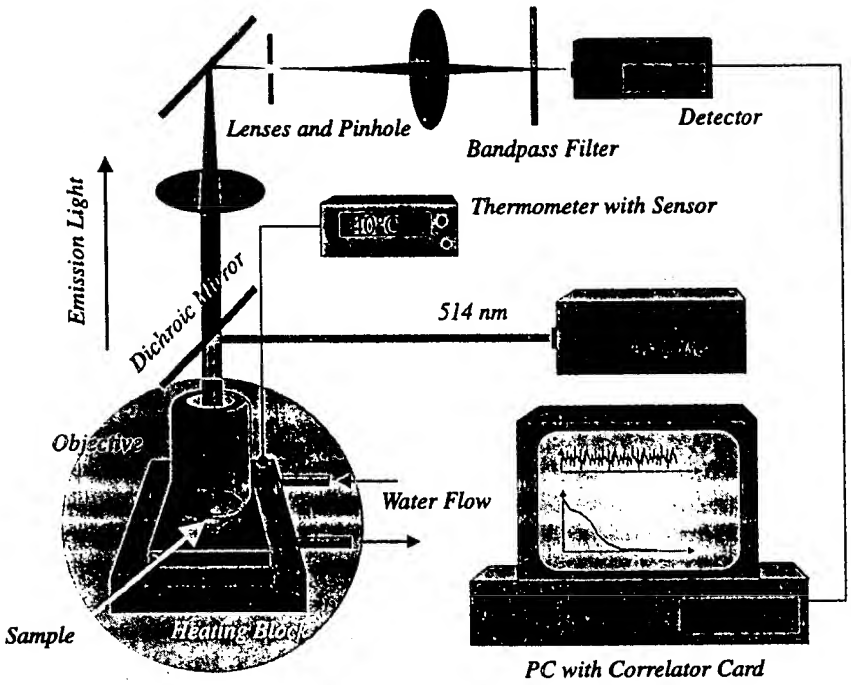

FIGURE 3: Schematic diagram of the fluorescence correlation spectroscopy (FCS) setup used in this study.

binding to target RNA was found to be negligible, as well as volume element instabilities due to temperature effects on the detection optics.

In practice, we had to include an additional diffusion time in the range of 0.01−0.04 ms to fit the autocorrelation curves of the labeled probes in the fast time range with satisfactory standard deviation. The fraction of this component was rather independent of laser intensity and slightly increased over incubation time at 40 °C. Therefore, it most likely represents either a very fast diffusing species like free fluorophore (which we did not detect by other means such as HPLC) or a bleaching term of a specific physical transition of TMR coupled to an oligonucleotide. Free probes and hybrid mixtures were evaluated by nonlinear least-squares fitting (Marquardt) of the obtained autocorrelation curves with eq 3 for $M = 2$ and $M = 3$, respectively. The diffusion time of the unknown fast component was calibrated tc 0.04 ms and held constant in the fits for all probes. Since this component was independent of RNA addition, its introduction allowed better fitting of the fast time range without affecting the calculated fractions of bound probe. To reduce the number of free fitting parameters and to clearly separate τ_i for free probe and hybrid, both were first determined independently by fitting the diffusion time of the probe without RNA and of the TMR-labeled RNA with excess unlabeled probe, respectively.

In Vitro Labeling of α-1 RNA. For fluorescent labeling of α-1 RNA, an *in vitro* transcription protocol (Milligan et al., 1987) was modified to include the TMR-labeled UTP of Figure 2 (TMR-12-UTP). The labeling reaction was carried out in a total volume of 500 μL for 1 h at 37 °C with 40 mM Tris-HCl, pH 8.0, 8 mM $MgCl_2$, 50 mM NaCl, 2 mM spermidine, 5 mM DTT, 1 mM each ATP, GTP, and CTP, 0.25 mM UTP, 0.125 mM TMR-12-UTP, 2 μg of *Hin*dIII-digested plasmid HP18α-1, and 10 units of T7 RNA polymerase/mL. The TMR-labeled transcript was purified by denaturing 7% PAGE and diffusion eluted, and its absorbances at 260 and 554 nm were determined. The DOS was calculated as described above to be 27%, indicating that a major fraction of fluorescent molecules carries a single TMR label while minor fractions carry two or more fluorophores.

Quantitated Primer Extension Assay. A 20 μL aliquot of a hybridization mixture was taken, supplemented with 3.5 μL of an assay mixture to give final concentrations of 1 mM of each dNTP and 0.53 units of (360 nM) of HIV-1 reverse transcriptase/mL and incubated at 40 °C for 2 min. Primer extension was stopped by adding 390 μL of a stop-mix containing 80 μL of water, 10 μL of 3 M NaOAc, pH 5.2, and 300 μL of EtOH. The labeled probe was precipitated by centrifugation, washed once with 70% EtOH, and dried, and half of it was loaded onto an 8% sequencing gel to be analyzed by electrophoresis on a model 373A DNA sequencer as described by the manufacturer (Applied Biosystems, Weiterstadt, Germany). After completion of the gel run, intensities of the fluorescent bands showing up in the yellow "T signal" were quantified, their relative distributions calculated, and their fragment lengths determined using the Genescan 672 equipment (Applied Biosystems, Weiterstadt, Germany).

DNA Melting Curves. Automated melting curves were recorded by monitoring A_{260} as described previously (Pörschke & Jung, 1982) using a Cary 219 spectrophotometer (Varian) on solutions containing the complementary DNA oligomers both at 5 μM in the same buffer as used in the hybridization protocols (60 mM Tris-HCl, pH 8.2, 10 mM $MgCl_2$, 10 mM KCl, 2.5 mM DTT, 2 mM spermidine). Temperature was increased from 10 to 90 °C, with a heating rate of 0.1 °C/min. Melting temperatures of the hybrids were determined by nonlinear least-squares fitting of the melting curves as described (Pörschke & Jung, 1982).

RESULTS

Following Hybridization with FCS. In our setup, the principle of fluorescence correlation analysis is combined with a confocal microscope (Figure 3). This allows to

autocorrelate temporal fluctuations in a very small volume element, restricted by the focal point of an epi-illuminated objective to about 0.2 fL. The beam waist of 0.5 μm is determined by the objective characteristics like numerical aperture and magnification, the five times larger z-dimension of the analyzed volume element is limited by a pinhole imaged in the focal plane. Both values proved to be constant during observation time in a previous measurement of calibrated pure dye solution of known concentration and diffusion properties. Temporal autocorrelation of the fluorescence signal from this illuminated open volume element yields information about characteristic diffusion times of the fluorophores. Since association of molecules results in higher molecular weights and increased diffusion times, hybridization can be followed online by a temporal decay shift of the FCS autocorrelation curve without separation of free and bound probe.

Hybridization of the six TMR-labeled probes HS1−HS6 to their target α-1 RNA (Figure 1) was typically performed at concentrations of 10 nM probe and 50 nM RNA to obtain kinetics with characteristic times in the 10 min range that could be followed over 1 h without special equipment for very fast reactions. Lower RNA concentrations resulted in kinetics too slow to be conveniently analyzed without the risk of RNA degradation. Since α-1 RNA contains part of the HIV-1 genome and has been used as target for *in vitro* reverse transcription using the viral polymerase (Pop, 1995; Gebinoga & Oehlenschläger, 1996), an HIV-1 reverse transcription buffer with 60 mM Tris-HCl, pH 8.2, 10 mM $MgCl_2$, 10 mM KCl, 2.5 mM DTT, and 2 mM spermidine at 40 °C was used as a typical environment for the underlying DNA−RNA hybridization reactions. Sonicated salmon sperm DNA at 10 μg/mL had to be added in order to suppress unspecific adsorption of probe and target nucleic acids at low concentrations to surfaces of the reaction chamber or microscope objective. Both FCS and primer extension analysis proved that salmon sperm DNA neither associated with probes nor with TMR-labeled target RNA. Under these conditions, FCS yielded autocorrelated fluorescence signals of the probes specifically shifting over time upon addition of complementary target α-1 RNA (Figure 4). No such shift was observed without α-1 RNA or after addition of nontarget strands like MDV-1 RNA (Mills et al., 1980).

To clearly separate diffusion times of free probe and hybrid, which only differ by a factor of 2−3, and to fix them in least-squares fits of the autocorrelation curves for better analysis, both were determined in independent measurements. For this purpose, labeled probe prior to addition of target RNA, and TMR-labeled α-1 RNA (generated by *in vitro* transcription in the presence of TMR-12-UTP, Figure 2) hybridized to a 50 times excess of unlabeled probe were analyzed, respectively. Diffusion times were calculated using eq 3 and are given in Table 1. The differences in diffusion times for the six hybrids might reflect the various extents of target secondary structure perturbation (Figure 1).

Fixing the obtained diffusion times of probe and hybrid in eq 3 enabled us to easily extract the distribution of the two fluorescent diffusing species from autocorrelation curves of the hybridization mixtures. However, especially for the first reaction phase (up to 40 min) the extracted diffusion times from fitting without fixing showed to be consistent with the calibration values (average errors of 3%−7%). Integration errors caused by a 30 s data collection time in the first reaction phase can be estimated to be below 5%. Figure 5 shows the increase in hybrid fraction over time for five of the probes. The observed kinetics are quite different, with HS1, HS5, and HS6 hybridizing rapidly and HS3 and HS4 being comparably slow. HS2 showed an increase in hybrid fraction over 1 h too low to be reproducibly quantified. It is obvious that, though a 5 times (in the case of 30 nM HS3, 1.7times) excess of target over probe was used, none of the probes quantitatively forms hybrids within the observation time. Generally, after a fast initial phase, the kinetics slow down such that after 1 h a considerable portion (typically between 10% and 40%, in the case of HS2 even up to 90%) of probe remains unhybridized. A limited hybridization extent was also observed for a different hybridization protocol, where probe and target RNA were denatured together and subsequently cooled down to rapidly obtain a maximum yield of hybrid (Table 2). In order to prove that this observation was not simply due to a lack of FCS to distinguish between free and bound probe or due to a detection bias for the faster diffusing free species, a quantifiable primer extension assay was designed as an independent measure for hybridization extent.

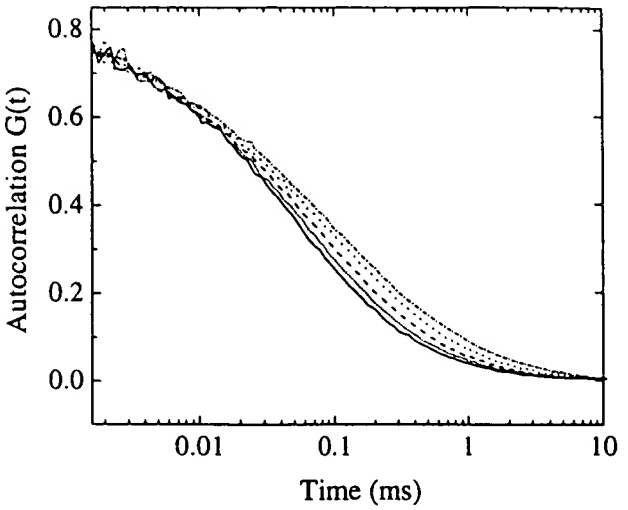

FIGURE 4: Shift over time of temporal autocorrelation $G(t)$ for 10 nM fluorescently labeled HS1 incubated with 50 nM α-1 RNA at 40 °C in hybridization buffer with 60 mM Tris-HCl, pH 8.2, 10 mM $MgCl_2$, 10 mM KCl, 2.5 mM DTT, 2 mM spermidine, and 10 μg of sonicated salmon sperm carrier DNA/mL. The half-value of the amplitude represents the average diffusion time. (Solid line, pure probe; short-dotted line, with RNA after 30 s; dashed line, after 5 min, dotted line, after 30 min; dash-dotted line, after 60 min)

Table 1: Diffusion Times of HS1−HS6, Free and Bound to α-1 RNA, through the Laser-Illuminated Open Volume Element of the FCS Setup in Hybridization Buffer at 40 °C (ms)

probe	HS1	HS2	HS3	HS4	HS5	HS6
free probe	0.15	0.18	0.21	0.20	0.11	0.15
bound probe	0.45	0.37	0.45	0.45	0.48	0.45

Comparison with Quantitated Primer Extension Assays. Since hybridization was performed in a reaction buffer for HIV-1 reverse transcriptase, the extent of hybridized TMR-labeled probe could easily be accessed by addition of this enzyme together with dNTPs at concentrations of 360 nM and 1 mM, respectively. The polymerase binds to DNA−RNA heteroduplexes with a binding constant of 5 nM (Kati et al., 1992), suggesting that the expected ≤10 nM hybrid in the hybridization assay should be readily bound by the

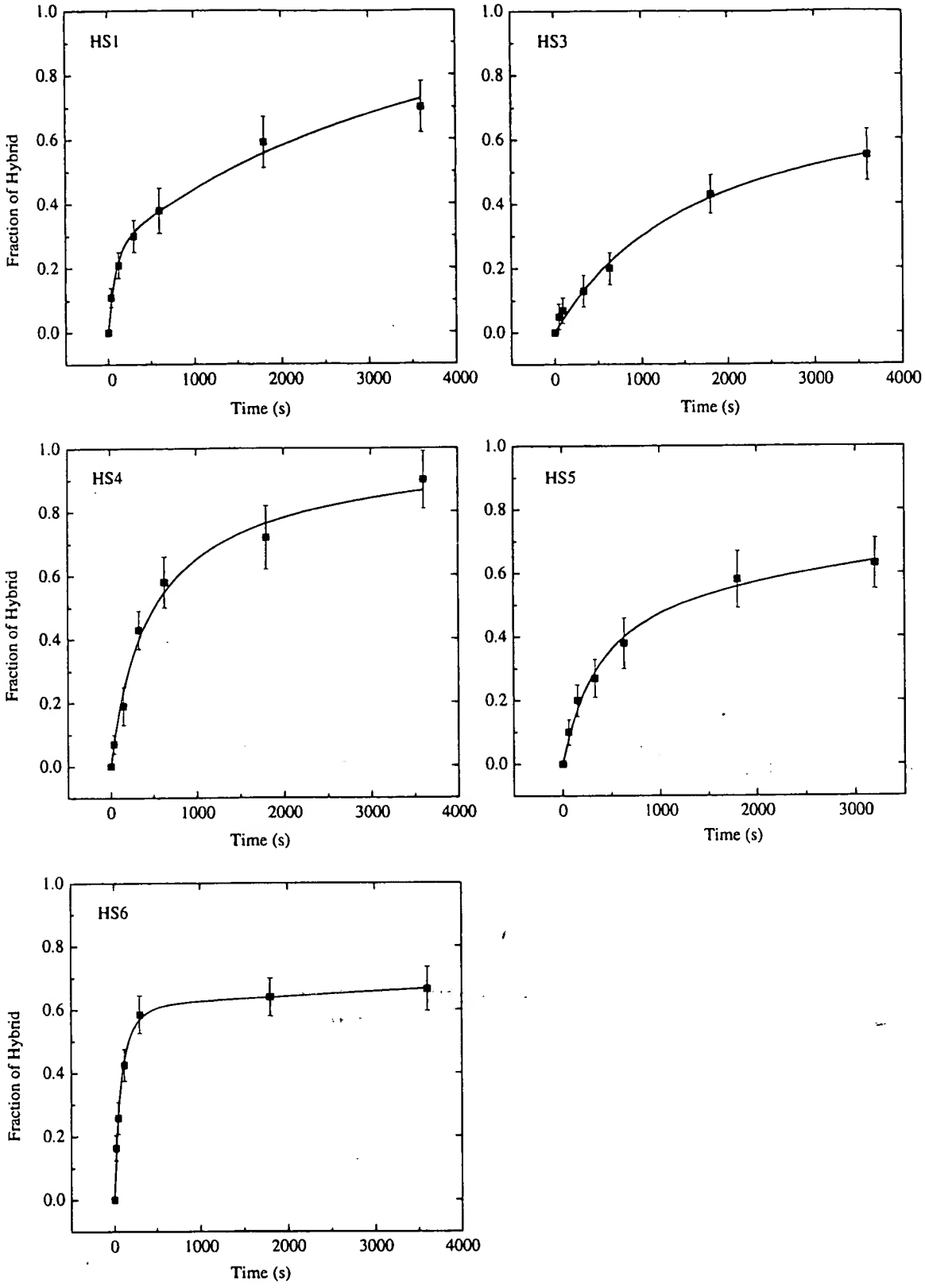

FIGURE 5: Hybridization kinetics of the probes HS1 and HS3−HS6 at 10 nM (30 nM in the case of HS3) with 50 nM α-1 RNA as measured by the shift in their autocorrelation function upon hybridization. Incubation buffer was 60 mM Tris-HCl, pH 8.2, 10 mM $MgCl_2$, 10mM KCl, 2.5 mM DTT, 2 mM spermidine, and 10 μg of sonicated salmon sperm carrier DNA/mL. Quantitative values for the bound probe fractions were calculated using eq 3 after determining the diffusion times for free probe and hybrid independently by analyzing the autocorrelation functions of probe without RNA and of TMR-labeled target with excess unlabeled probe. The solid line curves are fits obtained using eq 5.

enzyme. Moreover, HIV-1 reverse transcriptase incorporates nucleotides at a rate of 74 s^{-1} and dissociates from the DNA−RNA heteroduplex at 0.06 s^{-1} (Kati et al., 1992). Incubation for 2 min at a reaction temperature of 40 °C should therefore result in full extension of all probes hybridized to the 101mer RNA target, while free probe molecules should be unaffected. The obtained concentration of extension products was high enough to be analyzed and

Table 2: Hybridization Extent (with Standard Deviation of at Least Two Independent Measurements) after Incubation of Probe HS1−HS6 with α-1 RNA Target at 40 °C for 1 h and after Denaturation of Probe and Target Together[a]

probe	HS1	HS2	HS3	HS4	HS5	HS6
1 h at 40 °C/FCS	70 ± 6	*b*	55 ± 5	90 ± 9	60 ± 5	65 ± 5
1 h at 40 °C/E	*c*	11 ± 5	87 ± 5	89 ± 5	63 ± 5	78 ± 5
denaturation/FCS	65 ± 5	15 ± 8	40 ± 10	90 ± 5	65 ± 5	70 ± 5
denaturation/E	*c*	19 ± 5	64 ± 5	61 ± 5	54 ± 5	86 ± 5

[a] Binding fractions obtained by FCS are compared with results from the quantitated primer extension assay (E) (%), and all hybridization extents are in % of total probe; refer to Materials and Methods for detailed description of the two hybridization protocols. [b] Hybridization extent of HS2 was too low to be reliably measured by FCS. [c] HS1 binds to the target 5′ end and cannot be extended by reverse transcriptase.

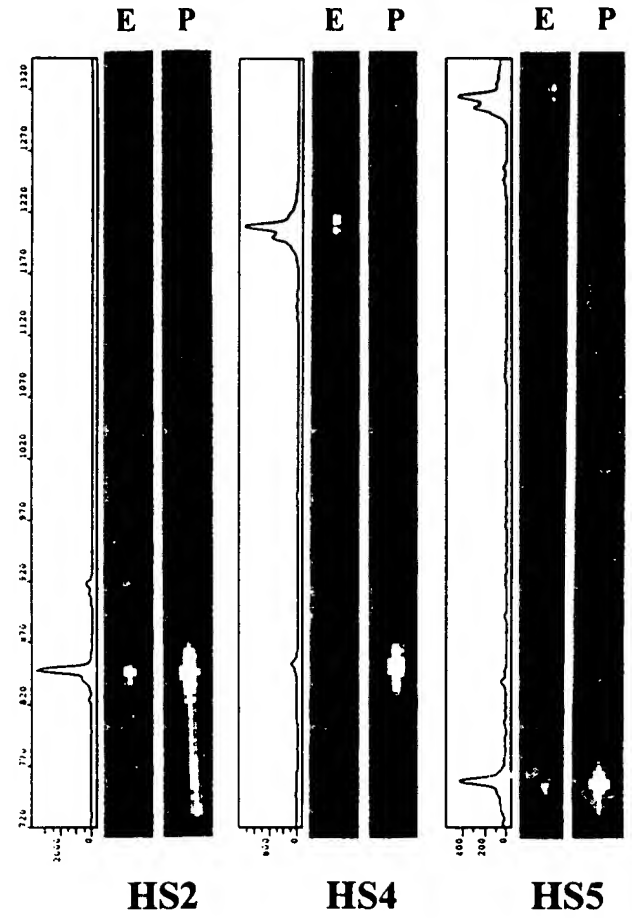

FIGURE 6: Principle of the applied quantitated primer extension assay with HS2, HS4, and HS5 as examples. After 1 h, samples from the hybridization mixtures were supplemented with dNTPs and HIV-1 reverse transcriptase, incubated over 2 min for elongation, the reactions stopped, the labeled probes precipitated and analyzed on a sequencing gel using the Genescan 672 equipment (Applied Biosystems, Weiterstadt, Germany). Lanes E were loaded with the extended probes, lanes P with the probes themselves. On the left of the lanes with extension products, the fluorescence scanning profiles are shown. HS2, HS4, and HS5 are probes showing about 10%, 90%, and 60% yield of extention product, respectively.

quantified after denaturing PAGE on an automated fluorescence sequencer. Figure 6 illustrates this novel quantitated primer extension technique. No elongation was observed without addition of target α-1 RNA, confirming the specificity of the reaction. All extension products were of the lengths expected for full extension to the target 5′ end, proving its integrity, with a characteristic double band indicating some 3′ end heterogenity, most probably due to incorporation of an additional nucleotide by the polymerase. Only between 10% and 90% of probe was elongated by HIV-1 reverse transcriptase during primer extension either after incubation with target at 40 °C for 1 h or after denaturation together with target and subsequent cooling down, with great differences between the six probes (Figure 6, Table 2). This did not essentially change with increasing duration of primer extension up to 20 min, confirming the results obtained by FCS.

Extraction of Kinetic Constants. The simplest way to interpret a limited hybridization extent as observed by FCS and primer extension assay even with target excess would be to assume a reversible hybridization reaction between probe and target with fast dissociation (Lima et al., 1992; Morrison & Stols, 1993). To have independent access to a dissociation rate constant, we tried to measure it directly by a method analogous to the label dilution method of Morrison and Stols (1993). Here, to a hybridized mixture of target and fluorescently labeled probe, a large (100 times) excess of the corresponding unlabeled probe HS1X−HS6X is added. Nevertheless, we did not find detectable dissociation for any of the probes.

By careful analysis of the hybridization kinetics of all probes, we found that they could best be described assuming a biphasic behavior with a fast initial and a slow second phase. The most simple process leading to such kinetics would imply the existence of the target RNA species RNA_f and RNA_s allowing fast and slow hybridization rates, respectively, and binding probe P with separable rate constants k_1 and k_2 to form hybrids $PRNA_f$ and $PRNA_s$, indistinguishable by FCS:

$$P + RNA_f \xrightarrow{k_1} PRNA_f$$

$$P + RNA_s \xrightarrow{k_2} PRNA_s \qquad (4)$$

For $k_1 \gg k_2$, this leads to the integrated rate equation

$$\frac{[PRNA]_{tot}}{[P]_0} = \frac{[PRNA_f]}{[P]_0} + \frac{[PRNA_s]}{[P]_0} = 1 - \frac{(1-m)}{1-me^{k_1P_0(m-1)t}} + \frac{(1-m)(1-e^{k_2P_0(1-\nu)t})}{1-\frac{(1-m)}{(\nu-m)}e^{k_2P_0(1-\nu)t}} \qquad (5)$$

with $[PRNA]_{tot}$ as the total concentration of hybrids $PRNA_f$ and $PRNA_s$, $[P]_0 \equiv P_0$, the initial probe concentration, m the ratio of initial RNA_f to initial probe concentration, $[RNA_f]_0/P_0$, and ν the ratio of total initial target to probe concentration, $[RNA]_0/P_0$, respectively. m gives a measure for the relative distribution of the two reaction paths. Fitting the observed kinetical hybridization curves with eq 5 yielded the solid curves of Figure 5. The three free fit parameters k_1, k_2, and m for the individual probes are listed in Table 3.

Direct Diffusional Analysis of Target α-1 RNA by FCS. In spite of thorough annealing of purified target α-1 RNA prior to analysis, non-denaturing PAGE as well as PACE indicated that several conformations of the RNA with different electrophoretic mobilities co-existed (data not

Table 3: Kinetic Constants k_1 and k_2 (M^{-1} s^{-1}) and Their Relative Distribution Parameter m for Hybridization of Probes HS1 and HS3−HS6 with α-1 RNA Target[a]

probe	HS1	HS3	HS4	HS5	HS6
k_1	$(1.3 \pm 0.4) \times 10^6$	$(3 \pm 1) \times 10^4$	$(3 \pm 1) \times 10^5$	$(3 \pm 0.7) \times 10^5$	$(1.5 \pm 0.5) \times 10^6$
k_2	4×10^3	*b*	4×10^3	3×10^3	1×10^3
m	0.29	0.71	0.82	0.45	0.61

[a] Probe HS2 showed a too low hybridization extent to be kinetically analyzed. $m = [RNA_f]/P_0$ and represents a measure for the distribution of the initial probe concentration P_0 on the reaction pathways with the two hypothetical RNA species RNA_f and RNA_s. [b] Probe HS3 showed a k_2 too low to be reliably measured.

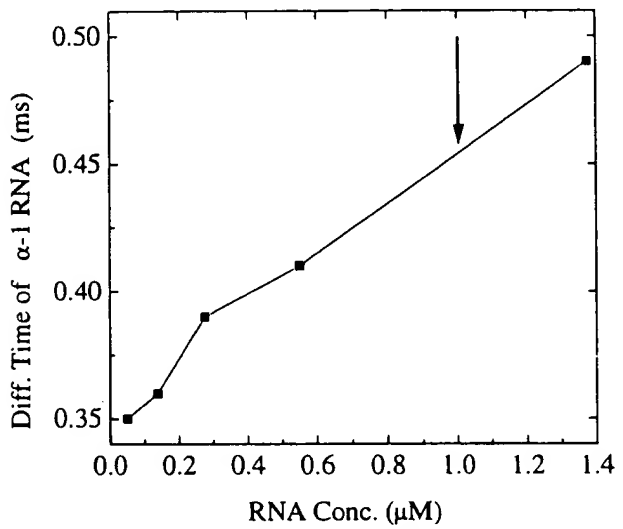

FIGURE 7: Diffusion time of target α-1 RNA in dependence of its concentration during initial denaturation in water. After cooling down, the RNA was diluted to a constant concentration in hybridization buffer including carrier DNA, and diffusion times were determined by FCS and analysis of the obtained autocorrelation curves. The standard initial RNA concentration prior to hybridization experiments was an intermediate value of 1 μM (arrow).

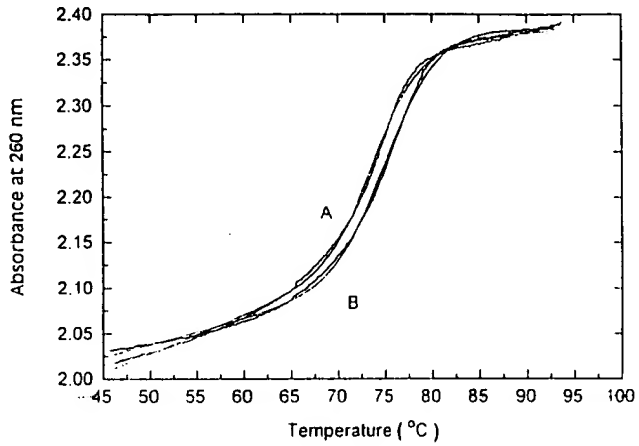

FIGURE 8: Melting curves of the labeled HS6-GSHS6 hybrid (A) in comparison with the unlabeled double-strand HS6X-GSHS6 (B, full lines). Both double-strands were measured at 5 μM in the same buffer as used in the hybridization protocols (60 mM Tris-HCl, pH 8.2, 10 mM $MgCl_2$, 10 mM KCl, 2.5 mM DTT, 2 mM spermidine). Also plotted are the best least-squares fits for the two experimental curves (smooth dotted lines). With the attached TMR label, the equilibrium melting point decreases from 73.5 (curve B) to 72.6 °C (curve A).

shown). This phenomenon has also been observed for other biologically important RNAs like group I introns at low (10 mM) $MgCl_2$ concentrations (Jaeger et al., 1991). In order to specify, whether oligomerization of α-1 RNA plays a role in forming these conformational inhomogenities, the RNA was heated in water at different concentrations, cooled down, and diluted into hybridization buffer, and the diffusion times in dependence of initial RNA concentration during denaturation were determined by FCS. Figure 7 shows that diffusion times nearly linearly increase over the examined α-1 RNA concentration range. Taking into account that diffusion times are roughly proportional to the third root of molecular weight of the diffusing species, the increase from 0.35 ms at 40 nM to 0.49 ms at 1.38 μM RNA would suggest the average formation of α-1 RNA monomers at 40 nM versus dimers or even higher oligomers by intermolecular hydrogen bonds at 1.38 μM, respectively. According to that, the presence of oligomers during the hybridization experiments must be taken into account. Since in free fitting, there was good consistency between observed complex diffusion times in calibration and hybridization measurements, the influence of probe binding to higher oligomers of RNA in the first reaction phase is shown to be of lower importance. However, it may be an explanation for the biphasic behavior, considering, e.g. RNA monomers as RNA_f and oligomers as RNA_s in eq 4.

Hybridization and Melting of TMR-Labeled DNA Double-Strands. Hybridization of TMR-labeled probes HS1−HS6 to an excess of their corresponding unlabeled cDNAs was performed as a control reaction. Table 4 indicates that for all six probes, a slight but significant increase in diffusion time upon hybridization could be observed. However, this increase was not high enough to clearly separate diffusion times of free probe and hybrid for quantitative FCS analysis of hybridization kinetics as described above for target α-1 RNA, thereby marking the limit for this kind of examination.

Table 4: Diffusion Times of Probes HS1−HS6, Free and Bound to Their Length-Matched cDNA GSHS1−GSHS6, through the Laser-Illuminated Open Volume Element of the FCS Setup in Hybridization Buffer at 40 °C (ms)

probe	HS1	HS2	HS3	HS4	HS5	HS6
free probe	0.15	0.18	0.21	0.20	0.11	0.15
bound probe	0.24	0.20	0.24	0.21	0.20	0.21

To study a possible influence of the TMR-label on hybridization of nucleic acids, the melting curve of TMR-labeled probe HS6 with its complementary unlabeled strand GSHS6 at equimolar concentrations in standard hybridization buffer was compared to that of the unlabeled HS6X/GSHS6 hybrid. Figure 8 illustrates that only a slight decrease of 0.9 °C in the presence of covalently attached TMR was found between melting points of the labeled and unlabeled double-strands.

DISCUSSION

In the present study, we used FCS-analyzed diffusion times to investigate binding kinetics of fluorescently labeled DNA probes to an artificial target RNA comprising part of the

HIV-1 genome as a model system for antisense oligonucleotide hybridization in solution. FCS proved to be a valuable tool for these kinds of studies because of (i) its high sensitivity down to the nanomolar concentration range, below which the kinetics become too slow for convenient analysis, (ii) the freedom to choose whatever buffer is desirable as long as it does not contain fluorescent contaminants, (iii) the possibility to follow hybridization in real time without separation steps for probe and hybrid, and (iv) the possibility to extract rate constants to quantitatively compare antisense oligonucleotides of interest, provided that their target is at least three to five times longer to significantly increase probe diffusion times upon hybridization.

We found, that the six probes HS1−HS6, being designed to have similar melting points with their target sequences, exhibit quite different initial association rate constants to α-1 RNA with k_i(HS6) > k_i(HS1) > k_i(HS5) ≈ k_i(HS4) > k_i(HS3) > k_i(HS2) (Figure 5; Table 3). These differences can plausibly be explained by different secondary structures of target sites and probes, thus confirming the predicted α-1 RNA structure (Figure 1). The fastest hybridizing probe is HS6, the one binding to the four non-base-paired nucleotides at the target 3′ end and to three internal loops. HS1 exhibits the second fastest association kinetics and hybridizes to six non-base-paired nucleotides at the target 5′ end and to one internal loop. HS5 is the third fastest probe. It binds to a stem−loop structure with one internal loop that represents the primer binding site of HIV-1 (though in the HIV-1 genome, the internal loop is part of a larger four-way junction). Obviously, this region is quite accessible for hybridization with an antisense sequence, a feature being essential for replication of HIV-1 (Isel et al., 1995). HS4 binds with the same rate constant as HS5, but to a higher extent. It is a long oligodeoxynucleotide hybridizing with its 3′ end to the largest internal loop of α-1 RNA and to three additional non-base-paired regions of the target. HS3 also is a long oligodeoxynucleotide and binds to the largest internal loop of the target, but both 3′ and 5′ ends of the probe bind to stems of α-1 RNA. Consequently, HS3 exhibits the second lowest rate constant. All five probes, HS1 and HS3−HS6, only show few base pairs in their own secondary structure prediction with largely single-stranded 3′ ends (Figure 1). Unlike these probes, HS2 as the least efficient binder has a tight stem−loop structure. This structure can interact with α-1 RNA only via a six-bases loop that is complementary to a similar structure in the target, while the rest of the target is hidden in a stem. These observations correspond very well with earlier studies on the pairing pathway of antisense nucleic acids, in which the binding process starts with a loop−loop interaction (called the "kissing complex"), and complete hybridization requires an additional interaction that involves single-stranded regions of both target and antisense nucleic acid (Siemering et al., 1994; Hjalt & Wagner, 1995). Consequently, the more internal loops and single-stranded regions are involved, the faster a hybridization will be (Lima et al., 1992), though tertiary interactions also play a role (Kumazawa et al., 1992; Zarrinkar & Williamson, 1994). Association rate constants of 10^6 M^{-1} s^{-1} as obtained for HS6 and HS1 in our study are thus among the highest known for antisense/RNA pairs.

As used in our study, FCS can only quantify overall hybridization kinetics by monitoring the distribution between free and stably bound probe. To account for the obtained complex kinetics with a fast initial and a slow second phase, for which we had evidence from both FCS and primer extension assays (Table 2), we had to assume a biphasic irreversible reaction. This is in contrast to an earlier study using FCS on an 18mer DNA probe hybridizing to a 7.5 kb long DNA target (Kinjo & Rigler, 1995), where monophasic irreversible kinetics were assumed, rapidly leading to 100% binding. Several reasons might account for this difference: (i) hybridization to a 400 times longer target results in a hybrid with a 20 times longer diffusion time, with little relative deviations masking small fractions of unbound probe; (ii) in the DNA−DNA hybridization study a single probe against a sequencing primer site was used, that can be expected to exhibit extraordinary fast and complete binding; (iii) hybridization of DNA to DNA is less complex than that of DNA to RNA due to fewer secondary and tertiary interactions within a DNA target. Our findings also contrast with other studies, where reversible hybridization kinetics have been proposed (Lima et al., 1992; Morrison & Stols, 1993). While we worked with rather long (17mer to 37mer) probes suitable as specific antisense agents and could not detect a significant dissociation from their target, in these earlier studies reversible hybridization was found for short (10mer) probes binding to short, low-structured DNA or RNA targets. Indeed, earlier studies on hybridization of long mRNAs with their cDNAs led to a complex multiphasic behavior similar to our results that was interpreted as a consequence of different sequence complexities of target sites (Young & Anderson, 1985).

A further hint for hybridization in our model system to be more complex than represented either by a single irreversible or reversible reaction of two species, is the fact that the relative contribution of the two reaction paths characterized by the ratio $m = [RNA_f]/P_0$ significantly differ between probes (Table 3). In this context it is necessary to keep in mind that both probe and target themselves are flexible structures that can co-exist in different secondary and tertiary structure conformations as found for other RNAs (Jaeger et al., 1991). In our system, evidence can be found from direct diffusional analysis of target α-1 RNA revealing different diffusing species in dependence of initial RNA concentration, an observation indicating multimolecular rather than unimolecular processes (Figure 7). Some of these conformations could lead to less accessible target sequences (Parkhurst & Parkhurst, 1995), resulting in separable phases in the overall reaction kinetics as the conformations react with different velocities. For probes with different target sites the contributions of different conformational species to the initial and second reaction phase can be expected to differ as experimentally observed. Thus, identification of different phases in hybridization as due to at least two different RNA conformations present in α-1 RNA preparations is a quite attractive interpretation of our results, though more complex reaction pathways resulting in biphasic behavior cannot be excluded. More detailed examination of the kinetics over a long time range could help to further elucidate the underlying reaction mechanisms. This will be a subject of further investigations. Since other kinetic interpretation models of limited hybridization extent, like the assumption of a totally unreactive RNA fraction, gave similar values for the initial association constants, the model applied here proved to be of lower importance for comparison of different probes with one another.

Antisense nucleic acids normally are not fluorescently labeled as the probes necessary for FCS analysis (Figure 2). We only found a subtle decrease of equilibrium melting points of our probes with length-matched cDNA upon coupling to the TMR label (Figure 8). Moreover, there is evidence that tetramethylrhodamine interacts with the nucleobases of an attached oligonucleotide (Bob Clegg, MPI, Göttingen, personal communication), so that an influence of the label on values of hybridization kinetic constants cannot be excluded. The relative comparison of kinetics of a set of antisense nucleic acids against a common target, however, should be unaffected by such an influence.

We therefore conclude that FCS is an appropriate tool for rapid screening for suitable antisense nucleic acids effective against targets of interest like HIV-1 RNA. After rapidly hybridizing probes were found, these could also be used to rapidly trace the presence of a sequence element by FCS, e.g., for detection of RNA by solution hybridization (Coutlee et al., 1990) or the automated analysis of mixed microbial populations in suspensions (Wallner et al., 1993). Moreover, by comparison of hybridization efficiencies of probes against different target regions, indirect evidence for predicted secondary structure elements might be possible. Using two probes simultaneously, one could identify higher-order structures by analyzing whether the binding of a fluorescently labeled probe is facilitated by preceding hybridization of an unlabeled probe to a distant target site. Easy synthetic or enzymatic access to suitable non-radioactive probes, low consumption of probe and target, free eligibility of additional probe modifications and hybridization buffer (even cellular extracts could be used) are important advantages of the method. We believe that this will extend the envisaged scope of fluorescence correlation spectroscopy applications (Eigen & Rigler, 1994; Rigler, 1995).

ACKNOWLEDGMENT

We thank Dr. Magda Pop for donation of α-1 RNA, Sylvia Völker and Dr. Günther Strunk for analysis of fluorescently labeled probes by HPLC, Dr. Dietmar Pörschke for measuring DNA melting curves, Dr. Said Modaressi for assistance with the GENESCAN Software, Prof. Rudolf Rigler for help in constructing the FCS setup, Prof. Fritz Eckstein and Dr. Franz-Josef Meyer-Almes for critical reading of the manuscript, and Prof. Manfred Eigen for a very stimulating environment.

APPENDIX

Deduction of the Fitting Function

Given the competitive reactions

$$RNA_f + P \xrightarrow{k_1} PRNA_f$$

$$RNA_s + P \xrightarrow{k_2} PRNA_s$$

with initial conditions for each component

fast: $[RNF_f]_0 \equiv RNA_f|_0 = m[P]_0 \equiv mP_0$

slow: $[RNF_s]_0 \equiv RNA_s|_0 = nP_0$

total: $[RNA]_{tot} = (m + n)P_0 = \nu P_0$

The relation for irreversible reactions is given by

$$[PRNA_s] = RNA_s|_0 - RNA_s|_0\left(1 - \frac{[PRNA_f]}{RNA_f|_0}\right)^{k_2/k_1} \quad (A1)$$

For $k_2 \ll k_1$ it follows that for the reaction of the fast component, $[PRNA_s] = 0$. This allows for small k_2/k_1 a separation in two time ranges.

first time range

$$\frac{d[PRNA_f]}{dt} = k_1[P](RNA_f|_0 - [PRNA_f])$$

$$\Rightarrow \frac{d[PRNA_f]}{dt} = k_1(P_0 - [PRNA_f])(RNA_f|_0 - [PRNA_f])$$

With $RNA_f|_0 = mP_0$ and $[PRNA_f] \equiv Y$

$$\frac{dY}{Y^2 - Y(m+1)P_0 + mP_0^2} = k_1 dt$$

With $X = ax^2 + bx + c$,

$$\int \frac{dx}{X} = \frac{1}{\sqrt{b^2 - 4ac}} \ln \frac{2ax + b - \sqrt{b^2 - 4ac}}{2ax + b + \sqrt{b^2 - 4ac}}$$

it follows that

$$\frac{1}{(m-1)P_0} \ln \frac{Y - P_0 m}{Y - P_0} = k_1 t + C$$

under the condition that

$$Y(0) = 0 \Rightarrow C = \frac{1}{(m-1)P_0} \ln m$$

$$\Rightarrow \frac{[PRNA_f]}{P_0}(t) = 1 - \frac{(1-m)}{1 - me^{k_1 P_0 (m-1)t}} \quad (A2)$$

second time range (the initial amount of probe is reduced by fraction of m)

$$\frac{d[PRNA_s]}{dt} = k_1(P_0(1-m) - [PRNA_s])(RNA_s|_0 - [PRNA_s])$$

with $RNA_s|_0 = (\nu - m)P_0$ and $[PRNAs] \equiv Y$

$$\Rightarrow \frac{dY}{Y^2 - Y(1 - 2m + \nu)P_0 + (1-m)(\nu - m)P_0^2} = k_2 dt$$

$$\Rightarrow \frac{1}{P_0(1-\nu)} \ln \frac{Y - P_0(1-m)}{Y - P_0(\nu - m)} = k_2 t + C$$

$$Y(0) = 0 \Rightarrow C = \frac{1}{P_0(1-\nu)} \ln \frac{(1-m)}{(\nu - m)}$$

$$\Rightarrow \frac{[PRNA_s]}{P_0} = \frac{(1-m)(1 - e^{k_2 P_0 (1-\nu)t})}{1 - \frac{(1-m)}{(\nu - m)} e^{k_2 P_0 (1-\nu)t}} \quad (A3)$$

The total reaction is given by the sum of eqs A2 and A3:

$$\frac{[\mathrm{PRNA}]_{tot}}{P_0} = \frac{[\mathrm{PRNA_f}]}{P_0} + \frac{[\mathrm{PRNA_s}]}{P_0} = 1 - \frac{(1-m)}{1-me^{k_1P_0(m-1)t}} + \frac{(1-m)(1-e^{k_2P_0(1-\nu)t})}{1-\frac{(1-m)}{(\nu-m)}e^{k_2P_0(1-\nu)t}}$$

REFERENCES

Bishop, J. O., Morton, J. G., Rosbash, M., & Richardson, M. (1974) *Nature 250*, 199−202.

Breslauer, K. J. (1986) in *Thermodynamic Data for Biochemistry and Biotechnology* (Hinz, H.-J., Ed.) pp 402−427, Springer-Verlag, New York.

Briggs, J., Elings, V. B., & Nicoli, D. F. (1981) *Science 212*, 1266−1267.

Britten, R. J., & Kohne, D. E. (1968) *Science 161*, 529−540.

Bush, C. A. (1974) in *Basic Principles in Nucleic Acid Chemistry* (Ts'o, P. O. P., Ed.) Vol. 2, pp 91−169, Academic Press, New York.

Coutlee, F., Rubalcaba, E. A., Viscidi, R. P., Gern, J. E., Murphy, P. A., & Lederman, H. M. (1990) *J. Biol. Chem. 265*, 11601−11604.

Crooke, S. T. (1992) *Annu. Rev. Pharmacol. Toxicol.* 32, 329−376.

Dewanjee, M. K., Ghafouripour, K., Kapadvanjwala, M., & Samy, A. T. (1994) *BioTechniques 16*, 844−850.

Dropulic, B., & Jeang, K. T. (1994) *Hum. Gene Ther. 5*, 927−939.

Ehrenberg, M., & Rigler, R. (1974) *J. Chem. Phys. 4*, 390−401.

Eigen, M., & Rigler, R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5740−5747.

Elson, M. E., & Magde, D. (1974) *Biopolymers 13*, 1−27.

Gebinoga, M., & Oehlenschläger, F. (1996) *Eur. J. Biochem. 235*, 256−261.

Hjalt, T. A. H., & Wagner, E. G. H. (1995) *Nucleic Acids Res. 23*, 580−587.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, L., & Schuster, P. (1994) *Monatsh. Chem. 125*, 167−188.

Inouye, M. (1988) *Gene 72*, 25−34.

Isel, C., Ehresmann, C., Keith, G., Ehresmann, B., & Marquet, R. (1995) *J. Mol. Biol. 247*, 236−250.

Jaeger, L., Westhof, E., & Michel, F. (1991) *J. Mol. Biol. 221*, 1153−1164.

Kati, W. M., Johnson, K. A., Jerva, L. F., & Anderson, K. S. (1992) *J. Biol. Chem. 267*, 25988−25997.

Kinjo, M., & Rigler, R. (1995) *Nucleic Acids Res. 23*, 1795−1799.

Koppel, D. (1974) *Phys. Rev. A 10*, 1938−1945.

Kumazawa, Y., Yokogawa, T., Tsurui, H., Miura, K., & Watanabe, K. (1992). *Nucleic Acids Res. 20*, 2223−2232.

Li, Y., Bevilacqua, P. C., Mathews, D., & Turner, D. H. (1995) *Biochemistry 34*, 14394−14399.

Lima, W. F., Monia, B. P., Ecker, D. J., & Freier, S. M. (1992) *Biochemistry 31*, 12055−12061.

Ma, C. K., Kolesnikow, T., Rayner, J. C., Simons, E. L., Yim, H., & Simons, R. W. (1994) *Mol. Microbiol. 14*, 1033−1047.

Magde, D., Elson, E. L., & Webb, W. W. (1972) *Phys. Rev. Lett. 29*, 705−708.

Magde, D., Elson, E. L., & Webb, W. W. (1974) *Biopolymers 13*, 29−61.

Manoharan, M., Tivel, K. L., Zhao, M., Nafisi, K., & Netzel, T. L. (1995) *J. Phys. Chem. 99*, 17461−17472.

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.

Mills, D. R., Kramer, F. R., Dobkin, C., Nishihara, T., & Cole, P. E. (1980) *Biochemistry 19*, 228−236.

Morrison, L. E., & Stols, L. M. (1993) *Biochemistry 32*, 3095−3104.

Müller, B., Restle, T., Weiss, S., Gautel, M., Sczakiel, G., & Goody, R. (1989) *J. Biol. Chem. 264*, 13975−13978.

Palmer, A. G., & Thompson, N. L. (1989) *Chem. Phys. Lipids 50*, 253−270.

Parkhurst, K. M., & Parkhurst, L. J. (1995) *Biochemistry 34*, 285−292.

Patel, D. J., Pardi, A., & Itakura, K. (1982) *Science 216*, 581−590.

Pop, M. P. (1995) *Impact of Sequence on HIV-1 Reverse Transcription*, Cuvillier Verlag, Göttingen, Germany.

Pörschke, D., & Jung, M. (1982) *Nucleic Acids Res. 10*, 6163−6176.

Qian, H., & Elson, E. L. (1991) *Appl. Opt. 30*, 1185−1195.

Rigler, R. (1995) *J. Biotechnol. 41*, 177−186.

Rigler, R., Widengren, J., & Mets, Ü. (1992) in *Fluorescence Spectroscopy* (Wolfbeis, O. S., Ed.) pp 13−21, Springer Verlag, Berlin.

Rigler, R., Mets, Ü., Widengren, J., & Kask, P. (1993) *Eur. Biophys. J. 22*, 169−175.

Siemering, K. R., Praszkier, J., & Pittard, A. J. (1994) *J. Bacteriol. 176*, 2677−2688.

Simons, R. W. (1988) *Gene 72*, 35−44.

Thompson, N. L. (1991) in *Topics in Fluorescence Spectroscopy* (Lakowicz, J. R., Ed.) Vol. 1, pp 337−378, Plenum Press, New York.

van der Krol, A. R., Mol, J. N. M., & Stuitje, A. R. (1988) *Gene 72*, 45−50.

Varani, G. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 379−404.

Wagner, E. G. H., & Simons, R. W. (1994) *Annu. Rev. Microbiol. 48*, 713−742.

Wagner, R. W. (1994) *Nature 372*, 333−335.

Wallner, G., Aman, R., & Beisker, W. (1993) *Cytometry 14*, 136−143.

Walter, N. G., & Strunk, G. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 7937−7941.

Widengren, J., Rigler, R., & Mets, Ü. (1994). *J. Fluorescence 4*, 255−258.

Widengren, J., Mets, Ü., & Rigler, R. (1995) *J. Phys. Chem. 99*, 13368−13379.

Yang, M. T., Scott, H. B., & Gardner, J. F. (1995) *J. Biol. Chem. 270*, 23330−23336.

Young, B. D., & Anderson, M. L. M. (1985) in *Nucleic Acid Hybridisation: A Practical Approach* (Hames, B. D., & Higgins, S. J., Eds.) pp 47−71, IRL Press, Oxford.

Zarrinkar, P. P., & Williamson, J. R. (1994) *Science 265*, 918−924.

BI960517G

EXHIBIT B

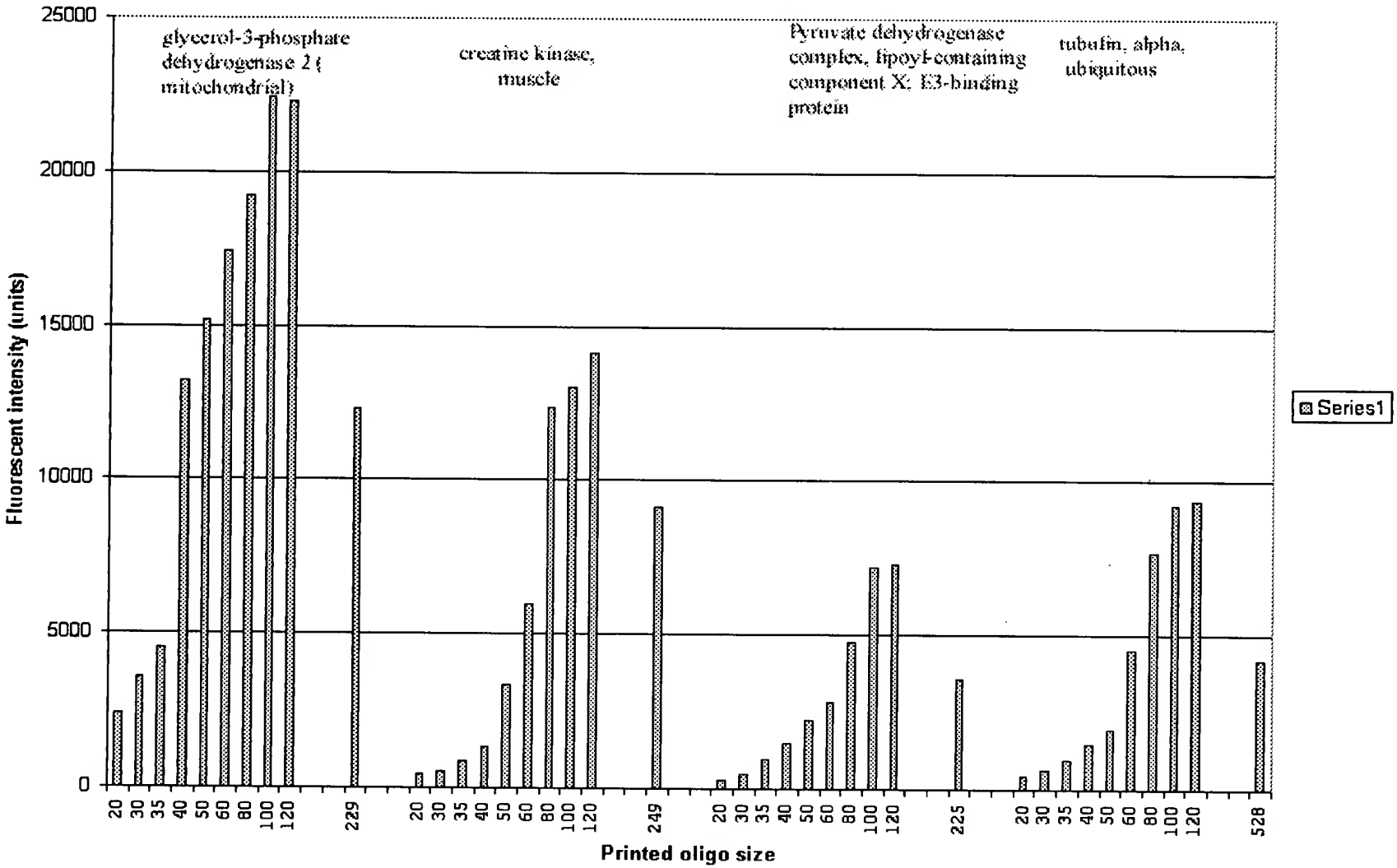
glycerol-3-phosphate dehydrogenase 2 (mitochondrial)
creatine kinase, muscle
Pyruvate dehydrogenase complex, lipoyl-containing component X; E3-binding protein
tubulin, alpha, ubiquitous
Fluorescent intensity (units)
25000
20000
15000
10000
5000
0
Printed oligo size
20 30 35 40 50 60 80 100 120 229
20 30 35 40 50 60 80 100 120 249
20 30 35 40 50 60 80 100 120 225
20 30 35 40 50 60 80 100 120 528
Series1

EXHIBIT C

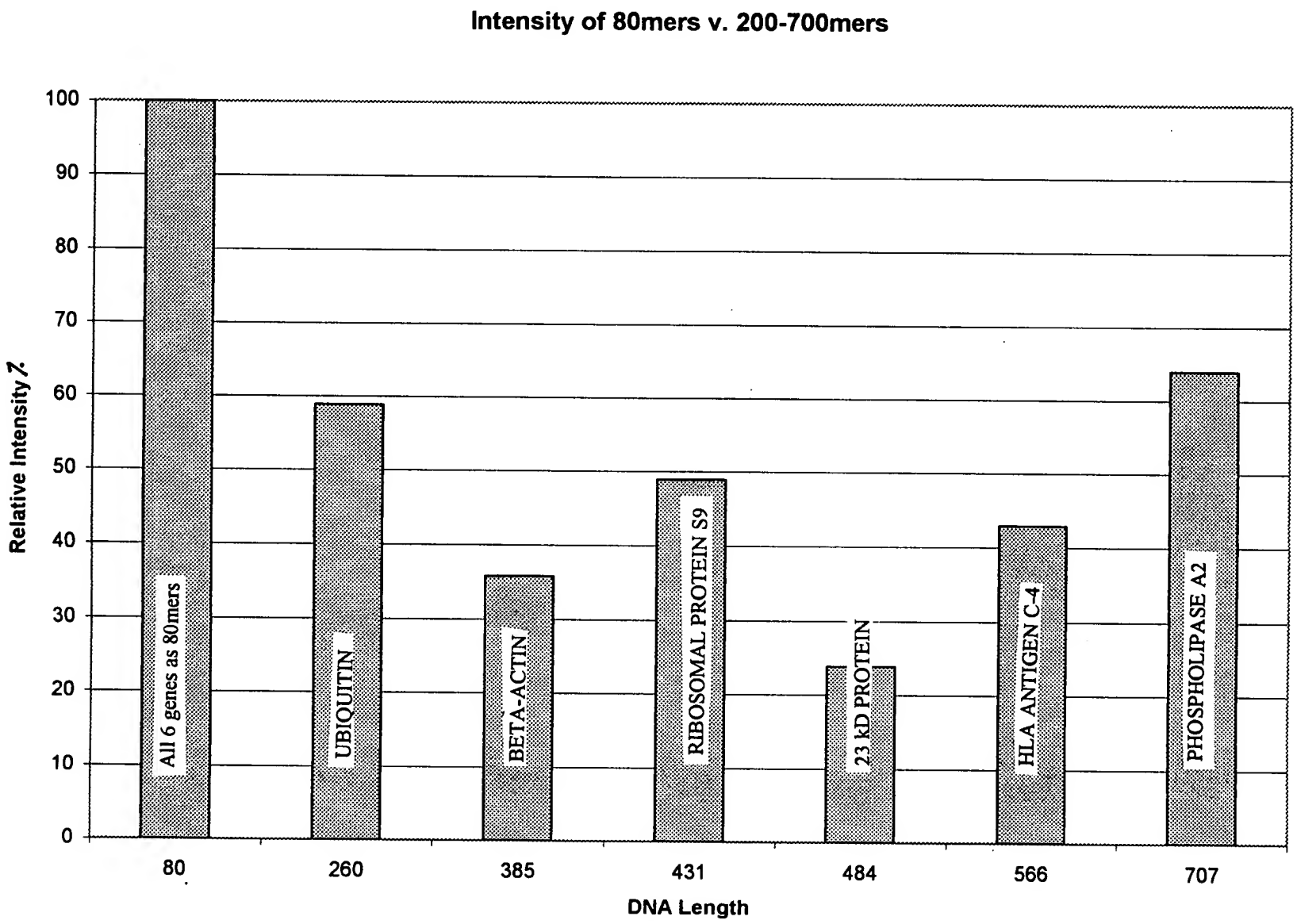
Intensity of 80mers v. 200-700mers
Relative Intensity %
100
90
80
70
60
50
40
30
20
10
0
All 6 genes as 80mers
UBIQUITIN
BETA-ACTIN
RIBOSOMAL PROTEIN S9
23 kD PROTEIN
HLA ANTIGEN C-4
PHOSPHOLIPASE A2
80
260
385
431
484
566
707
DNA Length

EXHIBIT D

Array-Ready Oligo Sets™

Array-Ready Oligo Sets™ consist of sets of optimized, arrayable 70mer oligonucleotides that can be used as probes for thousands of genes in a particular genome. Each 70mer is designed to have optimal specificity for its target gene and is melting-temperature normalized. Array-Ready Oligo Sets are available for human, mouse, rat, yeast, tuberculosis, malaria, Arabidopsis, C. elegans, and C. albicans. Custom array-optimized longmers (OPTs®) can also be designed by our bioinformatics team.

Features and benefits

- **Highly sensitive** — comparable to cDNA probes without the drawbacks of PCR amplification
- **Excellent specificity** — sequence optimized to minimize cross-hybridization, allowing analysis of overlapping and homologous genes
- **Normalized hybridization conditions** — melting-temperature (T_m) normalized oligo sets
- **Eukaryotic probes designed in the 3' region of each gene**
- **Ready to use**
- **Reliable and reproducible results**
- **Cost-effective alternative to cDNA- and PCR-based arrays**

Principle

Cross-hybridization between homologous DNA sequences is a long standing problem when analyzing gene expression using double-stranded DNA probes. For example, microarrays printed with PCR products may not allow detection of differential expression patterns for highly homologous or overlapping genes. In addition, PCR can be labor intensive and expensive, and has a high failure rate (typically 5–10%) when generating probes for microarray experiments (Figure 2.1).

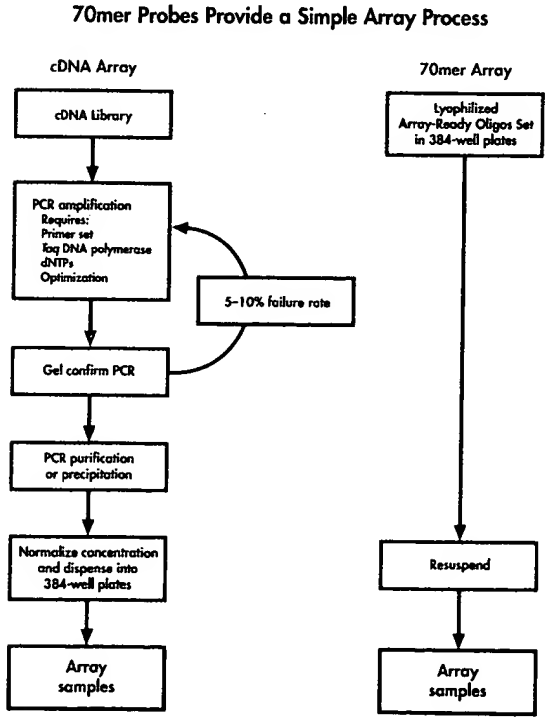

Figure 2.1 Comparison of the process for generating microarrays using cDNA and 70mer probes.

In contrast, synthetic DNA production is a robust and economical process for generating DNA probes for analysis of gene expression (Figure 2.1). We compared the sensitivity of 35, 50, 70, and 90mers for detecting highly expressed genes as well as genes expressed at moderate or low levels (Figure 2.2). 70mers were found to provide unique advantages, having the sensitivity of cDNA and the specificity of oligos.

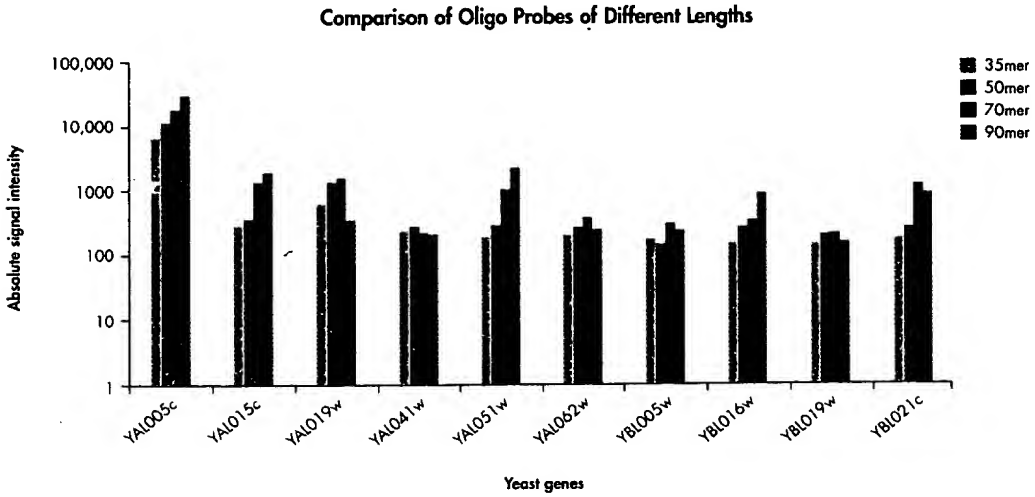

Figure 2.2 Microarray analysis of yeast gene expression levels using 35mer, 50mer, 70mer, and 90mer probes.

Probe design

Each Array-Ready Oligo Set contains 70mer probes representing thousands of well-characterized genes within the relevant genome. The sets were designed from genome databases by our bioinformatics team, often in collaboration with leading academic scientists. For eukaryotic genomes, each 70mer probe was designed with a bias towards the 3' end of the gene. The sequence of each probe was optimized using the search program Basic Local Alignment Search Tool (BLAST) by selecting the region of maximal specificity to the target gene while minimizing cross-hybridization to other genes. The 70mers are also designed to minimize secondary structure and are melting-temperature normalized to ensure consistent hybridization conditions. In the rare event that a gene-specific 70mer could not be designed, a shorter or longer oligonucleotide was designed, thus maintaining a consistent melting temperature (T_m). This sequence optimization procedure minimizes cross-hybridization and in many cases allows expression-level analysis of overlapping and homologous genes (Figure 2.3). In some very rare cases, gene sequences within a database are essentially identical, making it impossible to design unique probes for these sequences. In these cases, the probe sequences for different genes are identical.

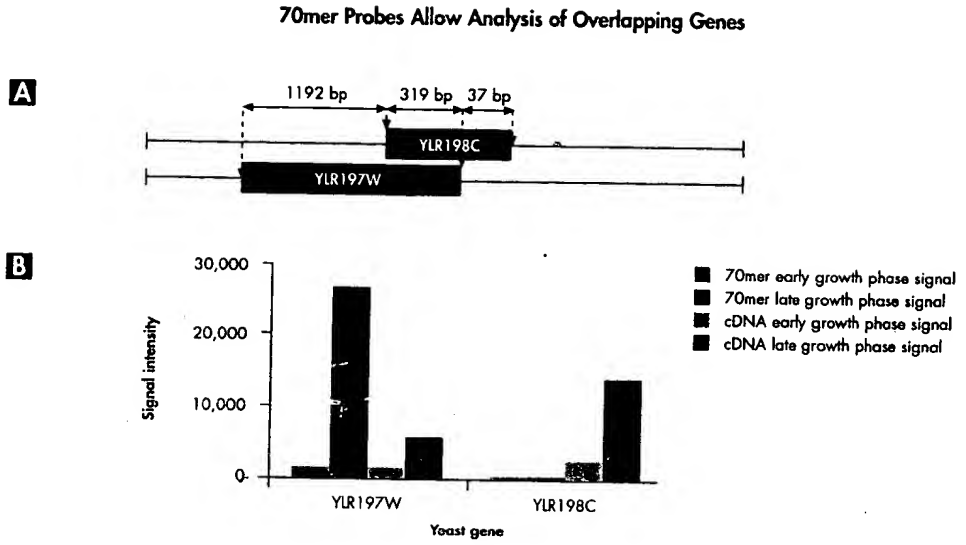

Figure 2.3 *Expression analysis of overlapping yeast genes.* **A** *The overlapping yeast genes* YLR198C *and* YLR197W *share 319 bp in their sequence.* YLR197W *is known to be activated as yeast shift from the early to the late growth phase, while the expression of* YLR198C *remains the same.* **B** *The expression of* YLR198C *and* YLR197W *during the early and late growth phases was examined using 70mer and cDNA probes. Both 70mer and cDNA probes showed that* YLR197W *expression is induced in the late growth phase, as expected. However, a* YLR198C*-cDNA probe showed activation of this gene in the late growth phase, a result not in agreement with published data. In contrast, the* YLR198C*-70mer probe gave the expected result.*

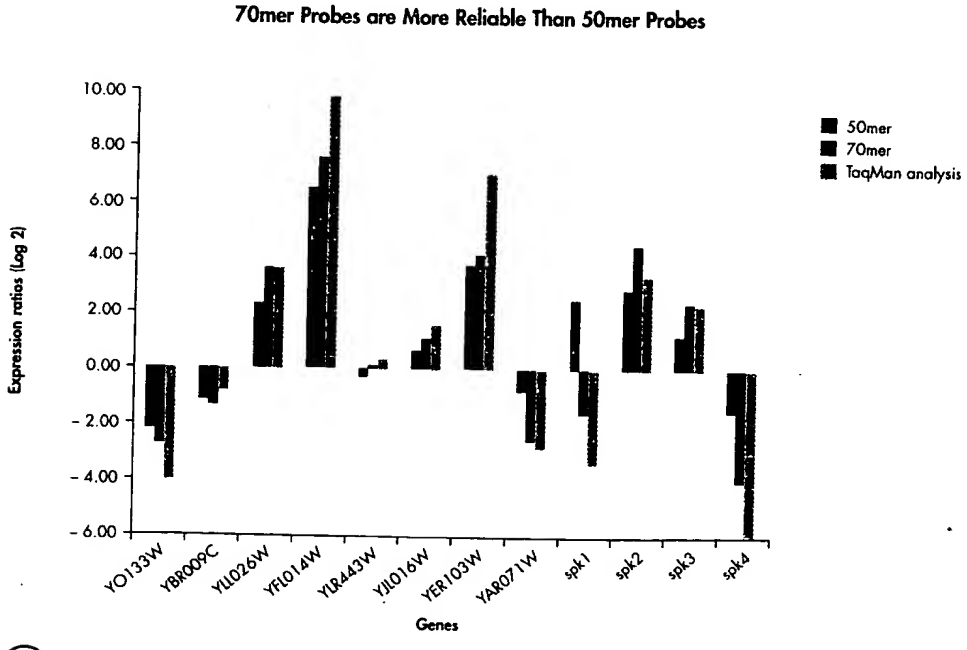

Figure 2.4 *Comparison of yeast gene expression levels using amino-linked 50mer and 70mer probes and TaqMan analysis. A strong correlation between the results from 70mer probes and TaqMan analysis was observed, demonstrating that 70mer probes are more reliable than 50mer probes for gene expression analysis. Probes were printed onto aldehyde slides. The mRNAs analyzed ranged in copy number from 0.4 to 23.9 copies/cell.*

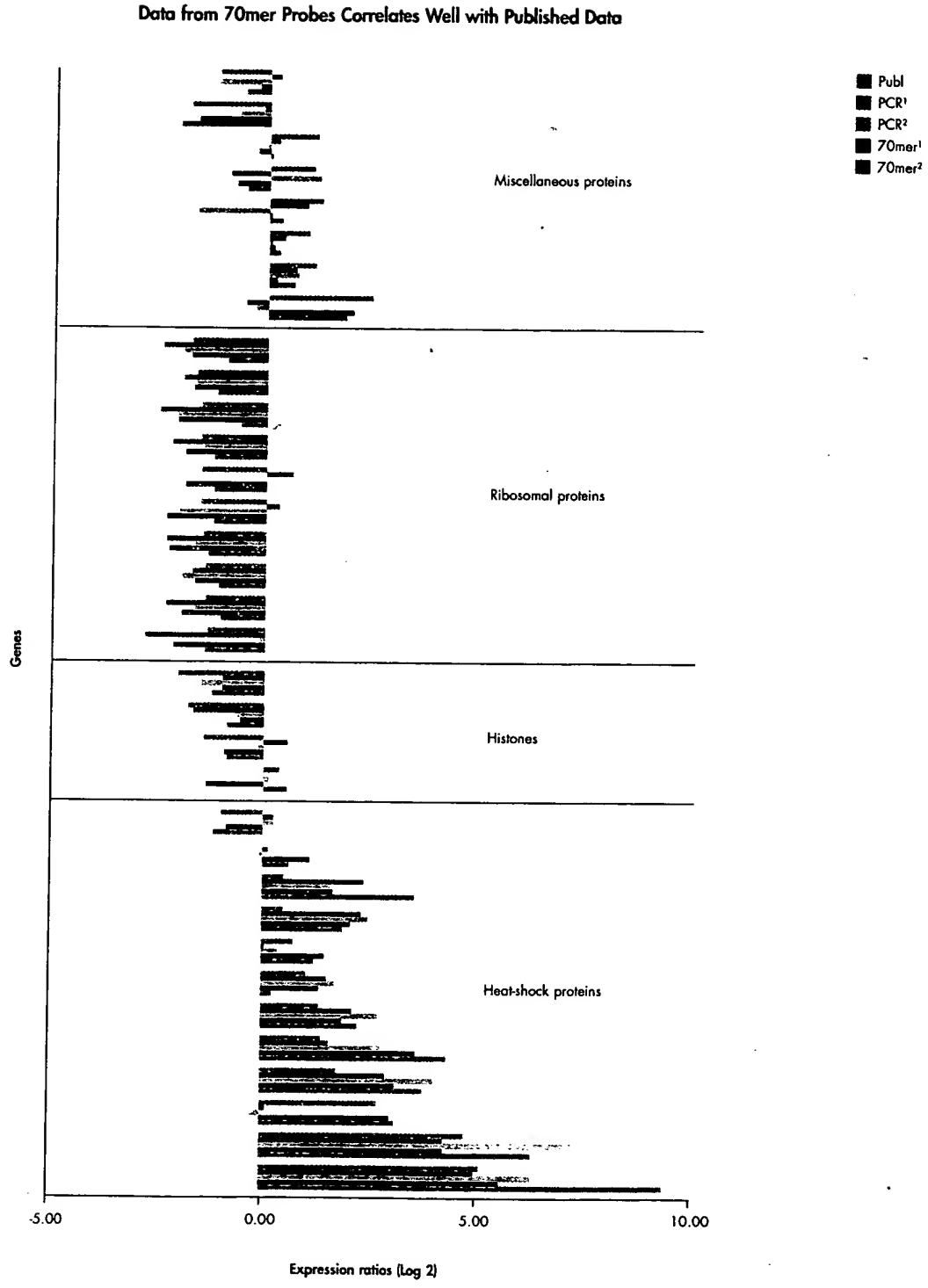

Figure 2.5 *Change in expression of yeast genes during heat-shock was analyzed on 2 microarrays printed with the Yeast Genome Oligo Set (70mer), 2 microarrays printed with PCR products representing whole ORFs (PCR), or by reference to published data (Publ; data from reference 1). Expression ratios are shown for specific gene families.*

Table 2.1 Correlation (*r*-value) of 70mer arrays, PCR-product arrays, and published data*

	70mer[1]	70mer[2]	PCR[1]	PCR[2]
70mer[2]	0.93			
PCR[1]	0.86	0.84		
PCR[2]	0.82	0.83	0.86	
Publ	0.89	0.89	0.80	0.76

** Calculated from Figure 2.5.*

Array-Ready Oligo Sets show reliable and reproducible performance in microarray experiments. The 70mer probes give more reliable data than 50mer probes (Figure 2.4). Compared to microarrays printed with PCR products, gene expression data from the sets show a higher correlation to published data (Figure 2.5 and Table 2.1). In addition, data from Array-Ready Oligo Sets are highly reproducible (Figures 2.7 and 2.8 and Table 2.1).

Analysis of *Plasmodium falciparum* Genes Using 70mer Probes

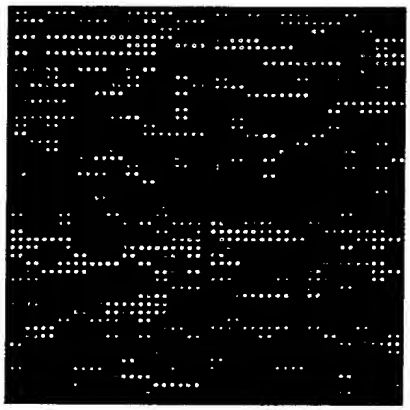

Figure 2.6 Microarray analysis using slides printed with 70mer probes from the Malaria Genome Oligo Set. A portion of the array is shown. (Data kindly provided by Dr. Rob Good, Walter & Eliza Hall Institute for Medical Research, Victoria, Australia.)

Formats

Array-Ready Oligo Sets are available for human, mouse, rat, yeast, tuberculosis, malaria (Figure 2.6), Arabidopsis, *C. elegans*, and *C. albicans* genomes. These sets are provided lyophilized in a 384-well plate format containing 600 pmol and 1200 pmol of each probe, which is sufficient for printing up to 1000 and 2000 slides, respectively. Please note that these slide numbers are approximate — the exact number will depend on the printing procedure used. The recommended printing concentration is 40 µM. Sample sets are also available containing 500 pmol of representative probes from each set.

Each Array-Ready Oligo Set is provided with a user manual and a computer disk containing detailed information on the set. The disk includes a list and description of the genes and controls included in the set, the plate position of each probe, and probe melting temperature information.

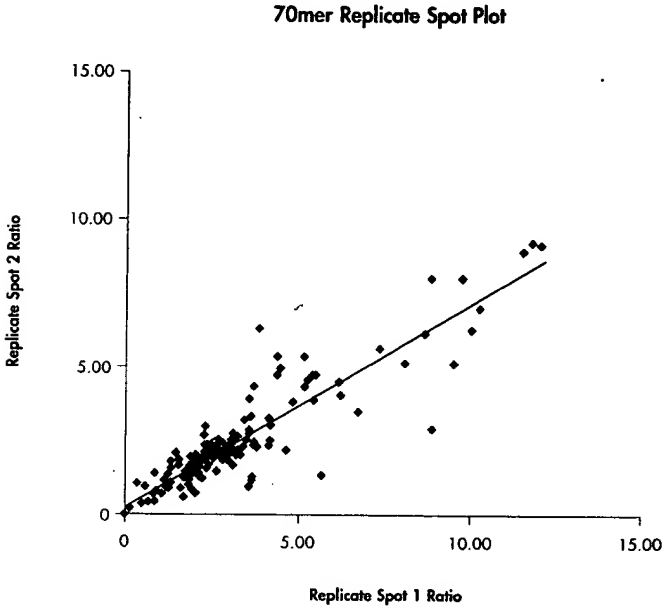

Figure 2.7 *Comparison of signal intensities from duplicate microarrays printed with 70mer probes for various yeast genes.*

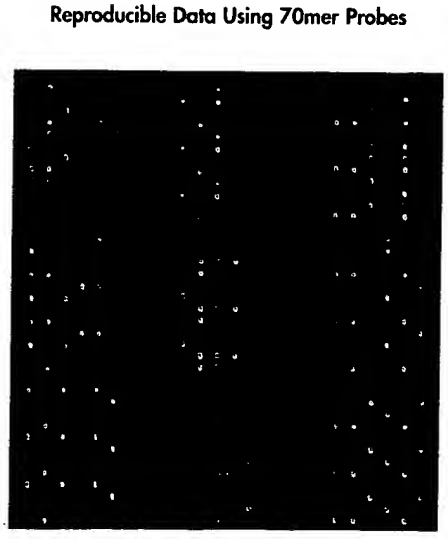

Figure 2.8 *Microarray analysis using slides printed with 70mer probes from the Human Genome Oligo Set. Each probe was printed in triplicate. A portion of the array is shown. (Data kindly provided by Dr. Shawn E. Holt, Virginia Commonwealth University, USA.)*

Web resources

Comprehensive technical information about each Array-Ready Oligo Set can be found on the QIAGEN Operon web site. The Operon Microarray Database (OMAD) (**www.operon.com/arrays/omad.php**) provides information on each probe (e.g., name, length, and melting temperature) and its representative open reading frame (ORF) or gene (e.g., name, symbol, and description) as well as links to public domain databases (Figure 2.9). Also included is potential cross-hybridization (X-Hyb) information based on sequence similarity of the probe to all other known genes or ORFs in the genome. This X-Hyb information can be used to provide a more in-depth analysis of the results of hybridization experiments.

Protocols for microarray analysis using Array-Ready Oligo Sets are available at **www.operon.com/arrays/arrayprotocols.php**. Protocols are provided for preparation of poly-L-lysine slides, printing, post-processing, QC 9mer hybridization, cDNA labeling, and hybridization. A troubleshooting guide is also provided.

Further information on Array-Ready Oligo Sets is also available from your local Technical Support Department.

Detailed Probe Information on OMAD

Oligo_ID	GB_accession	UniGene_ID	Gene_Symbol	Tm	Description
H000001_01	NM_012115	122843	CASP8AP2	73.6	CASP8 associated protein 2
H000002_01	AF035444	154036	TSSC3	78.8	tumor suppressing subtransferable candidate 3
H000003_01	AK001420	241531	PEF	77.7	peflin
H000004_01	M55150	73875	FAH	77.7	fumarylacetoacetate
H000005_01	AL121964	7510	MAP3K7	74.2	mitogen-activated protein kinase kinase kinase 7
H000006_01	NM_012094	31731	AOEB166	80.6	antioxidant enzyme B166
H000007_01	AK001917	80019	PDCD6	75.3	programmed cell death 6
H000009_01	U43342	248037	NFATC2	74.2	nuclear factor of activated T-cells, cytoplasmic 2
H000011_01	M13452	77886	LMNA	78.8	lamin A/C
H000012_01	AJ242832	225953	CAPN11	77.1	calpain 11

Figure 2.9 Example of 70mer probe and gene information available on OMAD. Probe length, plate position, and plate number are also provided on OMAD.

Cited References

1. *Lashkari, D.A. et al. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. USA* **94**, *13,057.*

Frequently Asked Questions

How do longmers compare to cDNA arrays?

The longmer approach to microarrays has been shown to provide better specificity (ability to differentiate between overlapping or related genes). QIAGEN Operon's bioinformatics group designs each oligo to have minimal homology with other human genes in order to reduce the chances of cross-hybridization. Additionally, the longmers are designed to have normalized melting temperatures closer to optimal hybridization temperature, again providing better specificity than cDNA arrays.

What kind of sensitivity do longmers provide?

Signal intensities of Array-Ready Oligo Sets under low stringency conditions are about 10% lower than those of cDNA arrays. The higher intensity cDNA signals are likely due to decreased specificity. cDNA arrays frequently report false positives due to cross-hybridization of related or overlapping genes.

What controls are included in the sets?

The Array-Ready Oligo Sets include negative and positive controls. The negative controls were chosen to have no homology to any other longmers on the array and should show no to very low signal intensity. The sets also include a positive control mix, an equimolar mix of all representative longmers in the original set. When printed, the positive control mix spot should show near-equal intensity from both Cy3 and Cy5. The difference in the intensity between the two dyes can be used to normalize the data for the whole slide.

Why is my negative control showing fluorescence?

The negative controls may exhibit some fluorescence due to cross-hybridization with a labeled probe. Although the control was designed to have minimal sequence identity, the BLAST alignment was limited to public domain databases, which may be incomplete. Cross-hybridization may occur as a result of an unknown or unavailable gene with sequence identity to the designed control.

Fluorescence from the buffer control spot may be a result of salt crystals on the slide that refract the laser light into the PMT. The fluorescence typically disappears after hybridization. Fluorescence from these controls after hybridization may be a result of carryover during the printing process.

How should the plates be stored?

The Array-Ready Oligo Sets should be stored lyophilized in –20°C.

Has anyone else used this product? Do you have a list of references who have used your product?

Many researchers both at academic and industrial institutions support QIAGEN Operon's longmer approach. A list of references is available upon request.

EXHIBIT E

50 nucleotide long probes on microarrays enable high signal intensity and high specificity

Dr. Susanne Schröder[1], Dr. Jaqueline Weber[2], and Dr. Hubert Paul[1]
MWG Biotech AG, Microarray Development[1] and Department of Bioinformatics[2],
Anzinger Str. 7, 85560 Ebersberg, Germany

Introduction

The signal intensity and specificity after hybridization of oligonucleotide-based microarrays depend mainly on the length of the spotted oligos and the hybridization conditions. A theoretical bioinformatic approach and a practical experiment were used to test the best suited length to achieve maximum signal intensity and specificity.

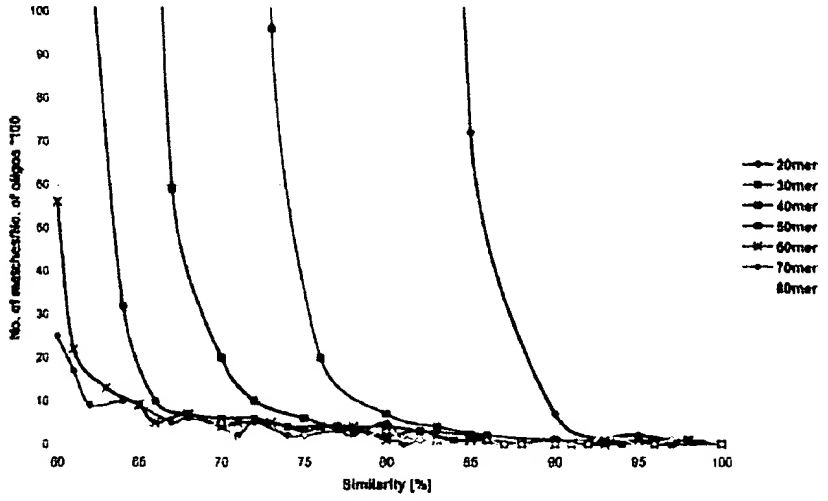

Figure 1
Bioinformatic estimation of the length-specificity relation. More than 75% similarity of a non-target position over the probe oligonucleotide length will lead to cross-hybridization. Thus, 20 and 30 mer oligonucleotide probes do not ensure a specific signal, whereas 50mer and longer oligonucleotides appear highly specific.

Bioinformatic estimation of the length-specificity relation

To determine the optimal length of the probe oligonucleotide non-overlapping oligonucleotides of different length were designed for 90 randomly chosen yeast genes, such that their sequences were 100% unique at their target positions. These oligonucleotide sequences were checked

MWGAG BIOTECH
meeting life's challenges

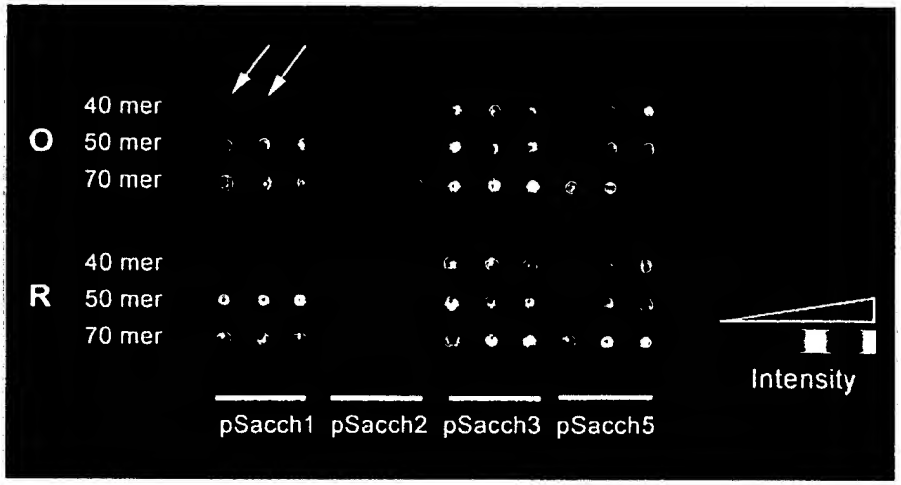

Figure 2
Array image after hybridization with cDNA of all four in vitro transcribed RNAs (pSacch 1, 2, 3, and 5). Two 40 mer oligonucleotides were replaced by negative control samples as indicated by arrows. In both the original (O) and the replicate (R) array the signal intensities increase the longer the spotted oligonucleotide is. Picture obtained by the Affymetrix 428 Scanner (available at MWG-Biotech AG).

against the whole genome of yeast for any similarity performing a Smith-Waterman alignment. It was analyzed how often a certain oligonucleotide sequence showing a particular degree of similarity can be found again in the yeast genome. Because the total number of short oligonucleotides, which can be designed for a given gene, will be higher than the total number of longer ones, more hits will be achieved.
For this reason, the number of matches with a certain degree of similarity was normalized to the number of oligonucleotides of a distinct length.

Specificity studies by KANE et al. [1] and the Microarray Development Department of MWG-Biotech show that for a given oligonucleotide probe any non-target transcripts

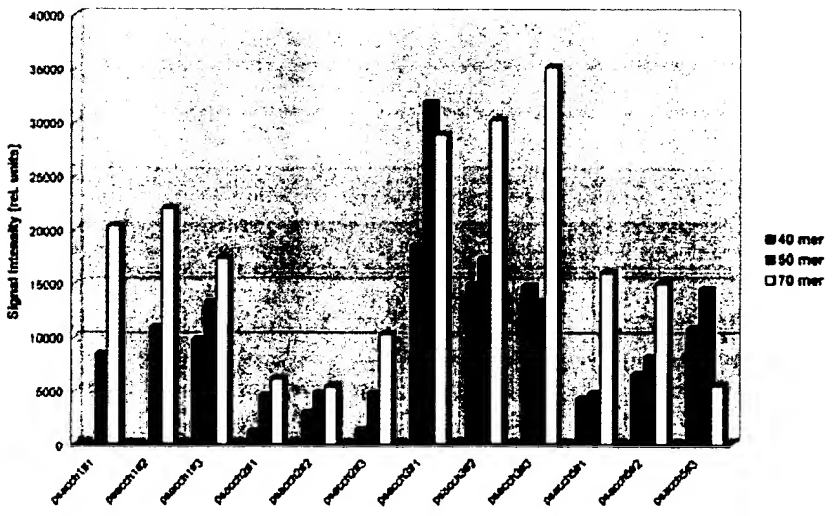

Figure 3
Diagram representing signal intensities depending on length of spotted oligonucleotides measured on the array in figure 2. Array analysis carried out using ImaGene Software, BioDiscovery (available at MWG-Biotech AG).

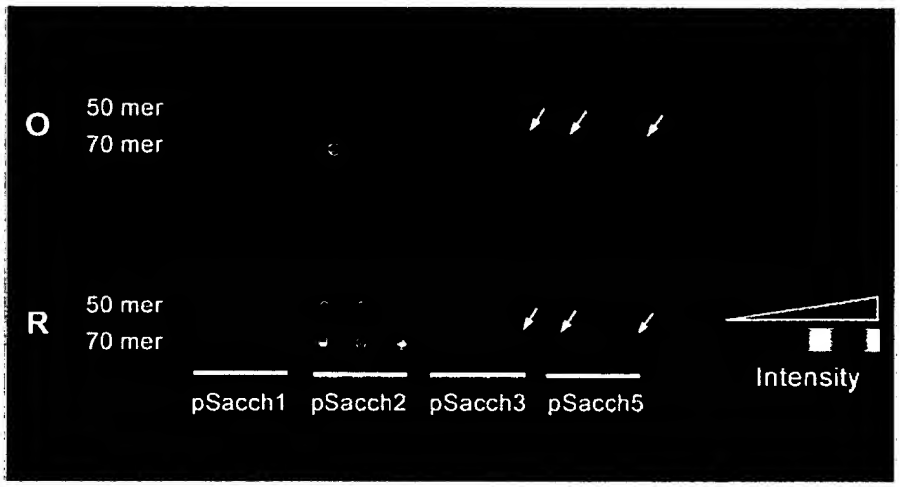

Figure 4
Array image after hybridization with cDNA of pSacch2 in vitro transcribed RNA. The hybridization results on the original (O) and the replicate (R) array show strong signals with the oligonucleotides targeting pSacch2, but also weak, unspecific signals with those targeting pSacch 3 and 5 as indicated by arrows. Picture obtained by the Affymetrix 428 Scanner (available at MWG-Biotech AG).

(cDNAs) with more than 75% similarity over the target length may show cross-hybridization. Non-target sequences which have >75-80% sequence similarity with target sequences within the oligonucleotide probe target region will contribute to the overall signal intensity. Thus, over its full length a specific probe oligonucleotide must not have more than 75% similarity to a non-target sequence.

Figure 1 clearly shows that for 20 and 30 mer oligonucleotides too many non-target sequences with a similarity >75% were found in the yeast genome to ensure a specific signal. Consequently, 20 mer and 30 mer oligonucleotides appear not suitable as probe oligonucleotides for microarrays. This figure also shows that 50 mer oligos are as specific as longer oligonucleotides. For both, 50 mer and longer oligos, very few non-target sequences with a similarity >75%

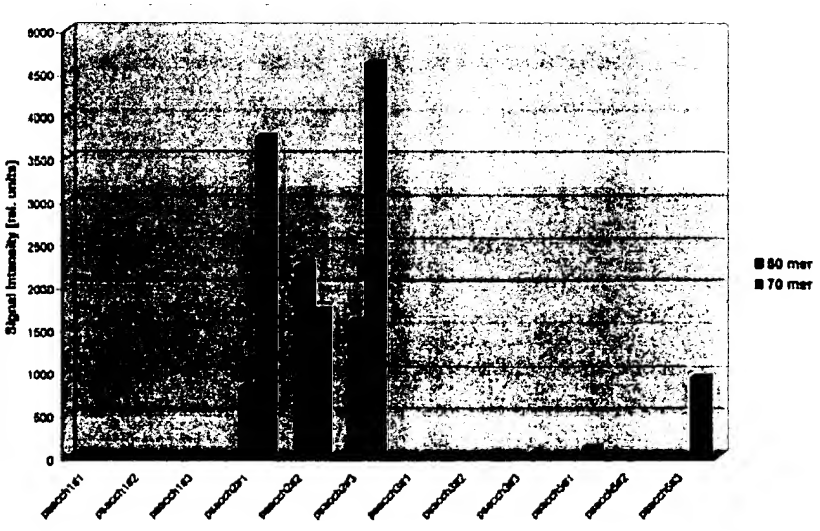

Figure 5
Diagram representing signal intensities of specific and cross-hybridization with cDNA of pSacch2 in vitro transcribed RNA measured on the array in figure 4. Array analysis carried out using ImaGene Software, BioDiscovery (available at MWG-Biotech AG).

were found in the yeast genome, proving that specificity cannot be increased by the use of longer oligonucleotides. Moreover, the longer the oligonucleotide is it becomes more likely to contain a stretch of complementary sequence of 15 and more contiguous bases. In only marginally similar targets, such a stretch may also lead to cross-hybridization [1]. The use of longer oligonucleotides will furthermore increase the possibility of secondary structure formation and will reduce the flexibility in oligonucleotide probe design.

Impact of length on signal intensity and hybridization specificity

To verify the results of the theoretical bioinformatic assay four different yeast genes (pSacch 1, 2, 3, and 5) were chosen and three oligonucleotides were designed for each of them targeting different positions in the gene under standard experimental conditions. These 12 oligonucleotide probes were spotted as 40, 50, and 70 mers on epoxy-coated slides generating an array of 36 spots. Afterwards they were hybridized in single and mixed assays with fluorescently labeled cDNA derived from 10ng in vitro transcribed RNA of the four yeast genes, respectively. This approach combining complex and specific hybridization allowed the precise detection of unspecific hybridization as well as direct comparison of signal intensities. Figures 2 and 3 show increasing signal intensities with increasing length of the oligonucleotides after hybridization with mixed cDNA of all four in vitro transcribed RNAs. Thus, in practice the transcripts may have easier access to longer oligonucleotides on the arrays, but unspecific binding may also contribute to the total signal, as the possibility of 15 mer stretches increases the longer the oligonucleotide is (as described above). However, it has to be taken into consideration that unspecific hybridization during experiments under standard conditions (42°C as incubation temperature, buffer containing formamide and SDS) is more likely the longer the spotted oligonucleotide is. Therefore, the array was also hybridized with cDNA derived from a single in vitro transcribed RNA, for instance pSacch2 (as shown in figure 4 and 5). As expected, the hybridization resulted in bright signals with the three pSacch2-specific oligonucleotides of each length, but more unspecific signals were observed with 70 mers of pSacch 3 and 5 oligonucleotides.

Conclusion

Combining bioinformatic and experimental results 50 mer probe oligonucleotides appear to be the best solution to achieve high signal intensities and high specificity on oligonucleotide-based microarrays.

[1] KANE, M. D. et al., Nucleic Acids Res. 28 (2000), 4552-4557.

For further technical information please contact:

GERMANY
Tel: +49-8092-8289 0
E-Mail: info@mwgdna.com
www.mwg-biotech.com

USA
Tel: +1-336-812-9995
E-Mail: info@mwgbiotech.com

ITALY
Tel: +39-055-4289-161
E-Mail: info@mwg-biotech.it

UK
Tel: +44-1908-525500
E-Mail: info@mwg.co.uk

SCANDINAVIA
Tel: +45-86172788
E-Mail: info@mwg.dk

FRANCE
Tel: +33-1-69592050
E-Mail: info@mwg-biotech.fr

MWGAG BIOTECH

meeting life's challenges

EXHIBIT F

bioArray News

THE GLOBAL WEEKLY OF BIOCHIPS & MICROARRAYS

volume 2, number 10
March 8, 2002

in this issue

Users Slow to Switch to Affymetrix U133s, but Initial Feedback is Positive

IT HAS BEEN over a month since Affymetrix marched out its latest U133 array set, with the whole draft assembly of the human genome squeezed onto just two chips.

But while the confetti and ticker tape has long been cleared from the parade route, few array users have used this new product enough to deliver a definitive verdict, according to an informal (and decidedly unscientific) survey of users conducted by *BioArray News*. Many have decided to finish all experiments with the previous-generation U95 human chips before getting started on the new arrays to avoid the technical headache of comparing results from the new chips to the old chips within the same experiment.

"The Affymetrix U133 arrays seem to be doing what they are supposed to. I think it's great that Affymetrix is giving you more genetic content for your dollar. But for us the jury is still out," said Andrew Brooks, director of the functional genomics center at the University of Rochester Medical Center.

NO NOISE IS GOOD NOISE

This lack of noise about the new chips could be seen as good news for the company, considering the past rumblings of activist

continued on page 4

Rosetta Researchers Use Microarrays to Define The 'Transcriptome,' Capture Splice Variants

ROSETTA Inpharmatics' $650 million acquisition by Merck last year hasn't stopped the company from conducting bleeding-edge microarray research. At the recent Microarray Gene Expression Database working group's annual meeting in Boston, Rosetta's Dan Shoemaker discussed the latest developments in the Kirkland, Wash., subsidiary's ongoing project to use oligonucleotide microarrays to discover new genes in the draft of the human genome and to monitor alternative splicing on a genome-wide level.

In the project, which Rosetta has previously described in a February 2001 *Nature* paper, Shoemaker and other Rosetta researchers are using microarrays to pinpoint more exactly the full library of transcripts in the genome by experimentally validating low-confidence predictions generated by *ab inito* gene-finding programs. To do this, the researchers

continued on page 5

PEOPLE IN THE NEWS

Former **Rockefeller University** president **Arnold Levine** will leave the university in June, according to Rockefeller sources. On February 11, the university announced that Levine had resigned as president, but would continue his research in cancer biology, in which he used microarrays to study mutations of the p53 gene in cancer.

As a result of Levine's impending departure, the university's core microarray facility budget has been scaled back, the sources said.

Luminex of Austin, Texas, has dismissed **Frank Reeves**, who served as chief financial officer since April 2001. The company will use an interim chief financial officer during the search for a permanent replacement. "There were no financial reporting issues that led to this", said **Mark Chandler**, Luminex's chairman and CEO. "The board just felt that going forward, they wanted to bring in somebody who had more public company experience in the accounting area."

NEW PRODUCT WATCH

Sorin Draghici and colleagues at **Wayne State University School of Medicine** have developed a software tool, **Onto-Express**, that maps differentially regulated genes in a microarray experiment to their function. The user enters into the program a list of accession numbers, UniGene cluster IDs or Affy probe IDs, Draghici said. The software then compiles a list that includes the functional categories associated with each gene, as well as statistics showing how many genes fit into each functional category. The categories include: biological process, cellular role, biochemical function, chromosome location; cellular component, and molecular function. This and other tools are available free at *http://vortex.cs.wayne.edu/Projects.html*

The European Bioinformatics Institute has released a beta version of MIAMExpress, a web-based submission tool for its public microarray database, ArrayExpress. The tool walks user through the requirements set by the Microarray Gene Expression Database working group, the minimum information about a microarray experiment, to ensure that submissions comply with this standard. EBI encourages users to practice using the tool, even though it is "not perfect yet," according to Alvis Brazma, team leader of the EBI's microarray informatics group at EBI.

ArrayExpress *(www.ebi.uk/microarray/ArrayExpress)* accepts three kinds of submissions — hybridization data sets, laboratory and data processing protocols, and array design.

CORRECTIONS:

In the 12/21/01 article "Could Microarrays Replace Sequencers? Callida Genomics Thinks So," Rade Drmanac, the chief scientific officer of Callida, was erroneously referred to as the chief executive officer. Also, Hyseq owns more than the 50 percent share of Callida indicated in the article, and Affymetrix provided an undisclosed sum to the company, not exactly $1.5 million.

In the 2/22/02 issue, the Patent Watch stated that Arcturus' RiboAmp RNA Amplification Kit sells for a list price of $1,125. This is out of date: Since November 2001, the kit has sold for a list price of $695.

bioArray News

The Global Weekly of Biochips & Microarrays

www.bioarraynews.com
ISSN 1534-9926

Marian Moser Jones, *Editor*
mjones@genomeweb.com

Julia Karow, DPhil, *Associate Editor*
jkarow@genomeweb.com

Gwen Libsohn, *Production Coordinator*
glibsohn@genomeweb.com

Marla Sansone, *Production Artist*
threebearsllc@aol.com

Dennis P. Waters, PhD, *Chairman*

Harry P. Greenwald,
Chief Operating Officer

Adrienne J. Burke
Editorial Director

SUBSCRIPTION INFORMATION

BioArray News is published weekly (50 times annually) by GenomeWeb LLC

Subscription rates:
North America ..... $845.
Outside No. America ..... €985.

To subscribe contact Miriam Smith
+1.212.651.5623
msmith@genomeweb.com

Or subscribe at our website at
http://www.bioarraynews.com

REPRINT ORDERS
Gwen Libsohn
+1.212.651.5630
glibsohn@genomeweb.com

ADVERTISING PLACEMENT
+1.212.269.4747
sales@genomeweb.com

GENOMEWEB, LLC
PO Box 998, Peck Slip Station
New York, NY 10272-0998 USA
Phone: +1.212.269.4747
Fax: +1.212.269.3686

Harvard Team Develops Absolute Expression Quantitation Method for Microarrays

AS MICROARRAY researchers know, relativity is a much less sexy concept in gene expression experiments than in theoretical physics. The ultimate aim would be to quantitate absolute levels of expressed transcripts. Until now, however, determination of absolute expression levels has been left to expensive bead-based technologies such as Lynx's Megasort system. But Aimee Dudley, a postdoctoral fellow, and John Aach, a senior scientist from George Church's Harvard Medical School lab, have developed a method for absolute quantitation of expression in yeast microarrays.

The traditional design for a microarray experiment involves hybridizing two-color labeled experimental and control samples to arrays, then examining relative expression levels between the two and adjusting for dye bias and other factors. But in this typical experiment, the absolute expression levels of transcripts in the control sample are unknown, so the final measurements of expression are just relative.

To get at this absolute quantitation, Dudley, Aach, and colleagues decided to replace the control sample with oligonucleotides that would be complementary to the yeast probe oligos on the array, ordering them from Qiagen Operon in small quantities. They divided the oligos into two samples, one of which was a five-fold dilution compared to the other; and labeled one sample with Cy3 and the other with Cy5, then hybridized them to the yeast array. In theory, their results would show a five-fold difference in expression levels, and given the known baseline concentration, they would be able to discern the expression level of an unknown sample. This turned out to be the case.

"This [technique] is directly applicable to comparing abundances between genes," said Dudley. In other words, by comparing the spot intensity for experimental sample (unknown) to that of the pre-measured oligos (known), they could derive the absolute number of transcripts of a particular gene expressed in the experimental sample.

The problem with this experiment was that the group was unable to order the 5,000-plus unique complementary oligos they would need to make up an entire reference sample for a microarray: Operon had not yet designed such a set.

PATENT WATCH

Protein chip startup **Nanotype** of Munich has received German Patent Number DE0010051143A1, "Process and Platform (or Apparatus) for the Characterization and/or the Detection of Binding Complexes".

The process the patent describes consists of the following steps: 1) Providing a first binding partner and a conjugate of a second and third binding partner, as well as provision of a fourth binding partner; 2) forming a linkage between the binding partners, where the first binding partner forms a probe complex with the second, and the third forms a reference complex with the fourth; 3) applying a force to the linked complexes which leads to the separation of the probe complex or the reference complex; and 4) determining which of the two binding complexes has been separated.

The patent broadly covers the company's C-Fit technology, which uses a nanoscale detection device, a nanotech version of an atomic force microscope, to detect forces between molecules as a method to detect binding between the probe and the target molecule. The patent's claims describe the entire platform as well as applications in nucleic acid, protein, and small-molecule assays.

Nanotype, which is currently collaborating with fellow Munich-based company **Xerion Pharmaceuticals** to develop a human nerve cell protein chip using its C-Fit platform, plans to secure patents in the US and other countries, and is also preparing to submit a paper to a scientific journal on the details of C-Fit, the company said.

The US Court of Appeals for the Federal Circuit has handed **Stratagene** a victory in an ongoing patent dispute sparked by **Invitrogen**, Stratagene said. The court's judgment, which upholds an earlier decision by a Maryland District Court, allows Stratagene to continue marketing its RNase H minus reverse transcriptase enzyme. Invitrogen filed suit against Stratagene on June 29, 2000, saying that its StrataScript RT RNase product infringed on Invitrogen's patent No. 6,063,608, and later asking the Maryland court to cease and desist its marketing efforts for the product, Stratagene said. Four months later, the court denied the request.

"The appellate Court's decision ensures Stratagene's ability to continue its rightful place in the RT market," the company said in a brief statement.

Invitrogen has also sued Strategene for infringement of patents related to its competent cell products.

So, to see if a similar concept would work on an entire array, they used defined quantities of complementary "universal" oligos that matched the universal primer tails of PCR products that had been spotted down on an array. They labeled the primer oligos with Cy3 and the RNA with Cy5, then were able to compare the expression level of the RNA to that of the primers.

This comparison of a known quantity of primers to an unknown quantity of transcripts not only allowed the researchers to quantify absolute expression levels, it also made it "easier to compare between experiments than if you use an unknown [quantity] of cDNA as your control sample," Dudley said.

This method additionally allows researchers to easily detect cross-hybridizations, or situations where oligos did not hybridize very well. These phenomena would be signalled by expression ratios that deviated from the five-fold dilution ratio in the first experiment.

The group has paired this new array system with "masliner," a program designed by Aach that extends the range of intensities that can be detected on an array, allowing an array user to use smaller amounts of oligos in each hybridization.

The group plans to publish a paper on this system in an upcoming issue of the *Proceedings of the National Academy of Sciences*, and meanwhile has found "a lot of enthusiasm" for it among other researchers, Dudley said.

"We think that companies like Operon will start producing mixtures of oligos complementary to their spotting oligo sets, once they become more familiar with this work and the interest of the microarray community in such a product," Dudley said.

— *MMJ*

Users Slow to...

continued from page 1

Affymetrix users. The company's disclosure of probe sequence on its NetAffx website — in response to longstanding criticism about the secrecy of these sequences — and its new algorithm seem to have muted criticism for the time being.

"I think [the new chips and sequences on NetAffx] are continuing to demonstrate Affymetrix's responsiveness to market forces," said Mark Geraci, who directs the microarray core at the University of Colorado at Boulder. "Just when you get frustrated with their product they do something that appeases the masses, and actually this one is quite thoughtful and quite helpful."

Those who have thoroughly test-driven the chips say they are pleased with the results. Gene Logic, which is the largest user of Affymetrix chips and runs through about 2,000 arrays per month, has delivered a glowing initial review. "We aren't finding that we have any problems with the U133 arrays. Technically they are running fine," Doug Dolginow, the company's senior vice president for pharmacogenomics, said last week. "It's clear that Affymetrix has really now moved into the commercial scale where these chips come out and tend to work because they are quality-controlled quite well, whereas four or five years ago there was not the same scale and reliability."

MINOR GRUMBLING OVER NETAFFX, ALGORITHM

Affymetrix user grumbling, however, has not stopped entirely. Some users complain that NetAffx is not user-friendly and has some rough edges. "I personally don't think that the design of the website is that fantastic — it's difficult to navigate — and for that reason I don't use it that much," said Brooks. Another criticism is that some U133 probes do not have annotations indicating what gene they represent. But these problems can be easily fixed and don't negate the site's usefulness, users agree.

"The ability to get at the oligo sequences for us has been a very big help," said Geraci. "But the site itself doesn't run entirely smoothly. It's just sort of a webmaster's problem."

Another grumble concerns the company's new data-mining algorithm. While many users have praised the company for delivering a statistically based improvement over the previous algorithm, "There are still problems with the new algorithm, and a number of people are contemplating writing their data extraction algorithms for Affymetrix chips," said Peter Tolias, director of the Center for Applied Genomics at the International Center for Public Health in Newark, NJ. "The new software removes the negative values when the mismatch hybridization is higher than the perfect match. They make it positive, and in doing so they mask the raw data. People want to know what the raw data is and make their own judgment as to whether a call is real or not," Tolias said.

HOW TO COMPARE U95, U133?

Complaints aside, users are chewing their pencil nubs over the problems involved in comparing data from the new chips and new algorithm to data generated with the old chips and old algorithm.

"People have to take a step back," said Brooks. "We have to

reevaluate how we assess the quality of our arrays based on the output being different as a function of a new array type and design, and the new algorithm." While NetAffx does ofter a visual comparison of the U95 and U133 probes, Brooks insisted that "you cannot actually truly compare the data between the two arrays yet" on NetAffx.

Mark Watson, director of Washington University-St. Louis' Affymetrix core facility, plans to resolve the comparison issue with "help from Affy and a bit of bioinformatics muscle. We won't compare directly between the two chips, but we will compare results of experiments run on U95s with new experiments run on U133s."

Others plan to do a test experiment on both chips, and then compare the data to see what kinds of changes there are on comparable probes. "I assume that Affymetrix must have done this experiment, but a lot of people are asking [about these results] and no one knows the answer," said Greg Khitrov, director of Rockefeller University's microarray facility.

Even with this comparison challenge, there is general agreement that Affymetrix' new offerings represent progress. And if imitation is really the sincerest form of flattery, Affymetrix has recently received a major compliment from Qiagen Operon. The company recently launched its own NetAffx-like web-based search tool, the Operon Microarray Database.

— MMJ

Rosetta Researchers...

continued from page 1

generated a comprehensive list of 116,000 transcripts by running multiple gene-finding algorithms on the draft of the human genome. The gene-finding algorithms were run at a very low stringency to capture as many new genes as possible. Unfortunately, casting this type of broad net also results in a large number of false positives, which is where the arrays come in, Shoemaker explained.

To determine which of the low-confidence predictions represented "real" transcripts, they designed four different 60-mer oligo probes for each of the 116,000 transcripts, and synthesized the resulting 500,000 or so probes on microarrays. They spotted the arrays using inkjet technology that Rosetta has licensed to Agilent. The group then hybridized the set of arrays covering all 116,000 transcripts with RNA from 60 different tissues to eliminate low-confidence transcripts with no expression activity. Based on the microarray experiments, they were able to narrow the list down to 68,000 transcripts. Shoemaker emphasized that this number is clearly an over-estimate because they have already found cases where multiple predictions that are next to each other in the genome can be collapsed into a single gene based on the co-regulation data from the 60 conditions.

One limitation of this approach, however, is that the gene finding algorithms have been trained on known genes and they might not be able to detect truly novel classes of transcripts. To identify these elusive transcripts, the group has tiled through 25 percent of the genome using 60-mers placed in 30 base-pair steps. They have hybridized the tiling arrays with RNA from six tissues, and the data is being used to experimentally determine the position of the exons in the genomic sequence.

In addition to refining the structure of known genes and discovering new genes, the Rosetta team is also using the tiling data to develop the next generation of gene-finding algorithms. Shoemaker said that their initial look at the data indicates that "there is a lot of transcription going on outside of annotated genes." The challenging task that lies ahead will be to determine what fraction of this transcriptional activity represents new genes vs. transcriptional noise.

The *Nature* paper covered findings for Chromsome 22, and now the researchers are looking to publish in a scientific journal their findings on all of the transcripts they have found in Chromosome 20, which represents about two percent of the human genome, Shoemaker said.

Shoemaker went on to describe how the Rosetta group is using microarrays to monitor gene structure — otherwise known as alternative splicing — on a genome-wide level. So far, they have mapped sequences of about 11,000 known RefSeq genes to the draft of the human genome to determine the position of the exons in each of the genes. This information was used to generate a set of "junction arrays," which were hybridized with RNA from 50 different tissues. Preliminary analysis suggests that approximately 30 percent of the genes had detectable levels of alternative splicing — some of which had not been previously reported in the literature. This type of finding has shown Rosetta researchers that "we need to monitor gene structure in disease-relevant tissues, and use co-regulation to infer biological function."

What will Merck do once the whole 'transcriptome' has been mapped out? The company isn't sure where to put the project, Shoemaker said, and he did not rule out the possibility that all of the results would be publicly released.

— MMJ

PROTEIN CHIP UPDATE: PEPTIDE CHIPS IN *NATURE BIOTECH*, ZYOMYX HOOKS UP WITH SPECIALTY LABS

The challenge of getting the right formula for surface chemistry on a chip has frustrated many would-be array makers in the protein arena. How do you immobilize the ligands without allowing nonspecific binding of proteins to the chip? Now a group of **University of Chicago** researchers has detailed, in an article in the March 2002 issue of *Nature Biotechnology*, a method for producing peptide chips that they claim overcomes these problems.

They used a clean gold substrate, which is immersed in alkanethiols 1 and 2 to make a monolayer. The glycol groups on this monolayer act to prevent nonspecific protein and radioisotope interactions, and the hydroquinone groups in the monolayer act as a "chemical handle" to selectively immobilize ligands, the researchers wrote.

The researchers tested this surface chemistry on a chip where they selectively immobilized a peptide-cyclopentadine conjugate of the peptide AclYGEFKKKC-NH2 on the monolayer, then used these chips to characterize the phosphorylation of the peptide by c-Src kinase. First they just used one type of peptide and one concentration of the kinase to test out the basic chemistry, and then they adapted the chip to characterize multiple kinase reactions, arraying mixtures of the c-SrC kinase with two kinase inhibitors, querceten and tyrphostin A47, at a range of concentrations.

For detection, the researchers tried out all of the three major techniques — radioactivity, fluorescence, and surface plasmon resonance spectroscopy, and found that the arrays worked with all three.

The group said it is now working on preparing arrays to identify protein-ligand complexes, to evaluate enzymatic activities such as protease and phosphate activities, and to "determine flanking sequences for enzyme substrates." They also suggested their chips could be useful for other small-molecule arrays.

Zyomyx, of Hayward, Calif., has granted clinical testing company **Specialty Laboratories** early access to its protein profiling biochip platform. Under the agreement, the two will work to discover multi-analyte diagnostic markers using Zyomyx's protein chips and Specialty's archive of biological samples. The two will also work together to design assay standards and controls, and Specialty will evaluate any assay the two companies develop. If Specialty does use a test, Zyomyx will retain exclusive rights to in-house testing services and will provide royalties to Zyomyx. Zyomyx will retain the right to develop and sell diagnostic test kits and therapeutic applications of these biomarker assays, and will pay royalties to Specialty, the companies said.

CARB CHIPS AREN'T JUST FOR SNACKING: COLUMBIA GROUP DEVELOPS CARBOHYDRATE ARRAY

A study published in *Nature Biotechnology* this month describes one of the first examples of a carbohydrate microarray. Researchers at **Columbia University** spotted 48 different carbohydrate antigens — including polysaccharides, glycosaminoglycans, glycoproteins, and semisynthetic glycoconjugates — onto glass slides coated with nitrocellulose, and air-dried them. To create the 150-micron-sized spots, 20,000 of which fit on a single microscope slide, they used a standard cDNA spot-arrayer. They found the arrays were very sensitive, enabling detection of sugar-binding antibodies in small samples of human serum. The scientists also used the glyco-chips to determine the specificities of known carbohydrate-binding antibodies and discovered unexpected crossreactivity, pointing to new epitopes for these antibodies *in vivo*.

In a preliminary experiment to test the binding of the carbohydrates — which does not involve chemical conjugation — to the surface, the scientists noticed that larger polysaccharides bound more efficiently than smaller ones, which limits the technology somewhat. Carbohydrates, which mediate molecular recognition between cells, have important roles in fertilization, development, cancer, and infection. Carbohydrate- or glyco-arrays may find widespread applications in these areas of research.

BIODISCOVERY AND GENOPS ENTER INTEGRATION PARTNERSHIP

BioDiscovery, based in Marina Del Ray, Calif., and **Genops Bioinformatics** of Vancouver, Canada, said last week that they had entered a strategic collaboration to merge their microarray- and sequence-based analysis platforms.

Under the terms of the agreement, BioDiscovery's GeneDirector, CloneTracker, ImaGene, and GeneSight data analysis products will be integrated with Genops' Ngene sequence analysis platform.

Financial terms of the agreement were not disclosed.

VBI's Rusty Kruzelock on Microarrays in New $10M Collaboration

Recently, the Virginia Bioinformatics Institute (VBI), and Johns Hopkins' Bloomberg School of Public Health began a five-year, $10 million collaborative bioinformatics project to study human infectious diseases such as malaria, tuberculosis, and HIV. Can you tell me a little bit about how microarrays will fit into the new collaborative project?

Microarrays will form the foundation for our studies. Of the total, the proportion earmarked for research is $285,000 for the first year, and about $177,500 for successive years.

AT A GLANCE

Research assistant professor, Virginia Bioinformatics Institute (VBI)

PhD in molecular genetics at the University of Texas, M.D. Anderson Cancer Center

Developed microarray technologies for Baylor College of Medicine and for Allan Bradley, director of the Sanger Center.

In the first year of the study, we will design a set of primers that can be used to make a genome-wide microarray. Oligos will be created using Primer3 software, along with homology-searching algorithms. These products will then be tested using VBI's real-time PCR platforms prior to array construction to ensure our predictions were correct. It's the same basic approach that has been used to create microarrays for microbes such as *E. coli*, *Helicobacter pylori*, and *Caulobacter crescentus*.

I understand you are going to develop microarrays to assess virulence factors of certain pathogens. Which pathogens are you going to select and why?

We are working closely with Johns Hopkins to produce microarrays to study malaria and tuberculosis. Pathogens have been selected based on their human health impact and modes of pathogenesis. For instance there are an estimated 300 to 500 million malaria cases each year, and increasing drug resistance is a major obstacle to effective treatment for both malaria and tuberculosis. Vaccines would clearly be a better control mechanism if we could understand how to make effective ones.

What sorts of arraying methods and equipment are you going to use to spot your arrays?

We use an epoxysilane attachment chemistry and spot the arrays with a GeneMachines Omnigrid using Telechem Stealth pins. For signal detection, we use a [Packard] ScanArray 5000XL. To store and analyze the expression data generated by our experiments, we will use GeneX, an open source gene expression database and analysis toolkit developed under the direction of Jennifer Weller, a scientist currently at VBI.

What are the major obstacles and challenges to using arrays?

As adequate controls and replicates for microarray data are fundamental requirements for producing quality results, more than 60 percent of our time is devoted to ensuring that adequate QC measures are put in place during the arraying process. Another significant challenge in dealing with arrays is managing the large quantities of data that result from experiments. There is currently an upload tool that makes it easy to insert Affymetrix data into the GeneX database, and a similar tool is under development to upload data from custom microarrays.

I have heard that microarray analysis presents some thorny problems. How do you analyze array data? What sorts of data mining techniques do you use?

Microarray data often contains considerable amounts of variation, stemming from both experimental and technical sources. The GeneX system includes a clustering procedure that takes this variation into account and assigns a confidence metric to all resulting gene groups. Implemented by Karen Schlauch at VBI, this tool supplies the user with a measure of "how sure" they can be that their clustering results are meaningful.

Also available within GeneX is a series of statistical tests that determine, at a specified significance level, whether genes are differentially expressed. Together with an ANOVA procedure to detect variation of specific effects, we believe that these tools will provide methods of generating meaningful hypotheses of the activity, interactions, and relevance of genes under infection.

What do you hope to gain from the array research?

We hope to develop an oligo-based microarray that can be printed in virtually unlimited quantities and shared by any faculty member interested in the physiology of the organism. We also are seeking to develop full-length expression competent constructs to assess the functional roles of genes critical in host-vector and host-pathogen infection processes; a vector system to generate site-directed or systematic loss-of-function mutations in host, vector, or pathogen genomes; and interesting and meaningful hypotheses of gene activity under infection conditions.